

# Chemistry and Medicine

PAPERS PRESENTED AT THE FIFTIETH ANNIVERSARY  
OF THE FOUNDING OF THE MEDICAL SCHOOL  
OF THE UNIVERSITY OF MINNESOTA

*Edited by*

Maurice B. Visscher

PROFESSOR OF PHYSIOLOGY AT THE  
UNIVERSITY OF MINNESOTA

THE UNIVERSITY OF MINNESOTA PRESS

*Minneapolis*

International Book House Ltd.,

**Bombay**

*Copyright 1940 by the*  
**UNIVERSITY OF MINNESOTA**

ALL RIGHTS RESERVED. NO PART OF THIS BOOK  
MAY BE REPRODUCED IN ANY FORM WITHOUT  
THE WRITTEN PERMISSION OF THE PUBLISHER.  
PERMISSION IS HEREBY GRANTED TO REVIEWERS  
TO QUOTE BRIEF PASSAGES IN A REVIEW TO BE  
PRINTED IN A MAGAZINE OR NEWSPAPER.

*Printed at The Lund Press, Inc., Minneapolis*

## FOREWORD

THE Committee of the Medical Faculty which was appointed to formulate an appropriate program for the commemoration of the fiftieth anniversary of the establishment of the Medical School of the University of Minnesota decided to focus the scientific program for this occasion upon the single theme: Some Trends in Medical Progress with Particular Reference to Chemistry in Medicine. This subject was chosen not because it is more important than other fields of medicine but because it represents one of the most recently developed and more rapidly expanding aspects of medical science.

The papers which were presented on the program of this fiftieth anniversary celebration (October 12, 13, and 14, 1939) were of such excellence that the university felt obligated to publish them in order that they might be permanently preserved and available both to those who attended the sessions and those who were unable to be present. The Anniversary Committee on Publications, under the chairmanship of Dr. Maurice B. Visscher, planned and arranged with the University of Minnesota Press for the publication of this volume.

We are gratified with the record that the Medical School of the University of Minnesota has made over the first half century of its existence and we accept the splendid scientific papers presented in this volume as a challenge to even greater achievement during the half century which lies ahead.

Harold S. Diehl  
*Dean of the Medical Sciences*  
*University of Minnesota*





## CONTENTS

### *Part I. Progress in the Application of Physical Chemistry to Medicine*

Some Aspects of the Colloid Chemistry of Membranes by <i>Herbert Freundlich</i> .....	3
The Performance of Osmotic Work in Living Systems by <i>Maurice B. Visscher</i> .....	16
Some Reactions by Which Solutes May Be Differentially Concentrated by the Kidney by <i>John P. Peters</i> .....	30

### *Part II. Some Recent Investigations in Metabolism*

Organic Chemistry in the Pursuit of Vitamin Research by <i>Lee Irvin Smith</i> .....	57
On the Necessity of Fats in the Diet by <i>George O. Burr</i>	101
Heparin and Thrombosis by <i>Charles H. Best</i> .....	121

### *Part III. Some Aspects of Immunity and Chemotherapy*

Recent Chemical Trends in the Study of Immunity by <i>Michael Heidelberger</i> .....	139
The Biology of Animal Viruses by <i>Robert G. Green</i> ....	157
The Mode of Action of Sulfanilamide and Its Derivatives by <i>Perrin H. Long</i> .....	185
Chemistry in Urinary Antisepsis by <i>Henry F. Helmholz</i>	212

### *Part IV. Some Approaches to the Nervous Control of the Organism*

The Physicochemical Approach to the Mechanisms of Convulsive Reactivity by <i>Irvine McQuarrie</i> .....	225
Methods of Analysis of Nervous Action by <i>Herbert S. Gasser</i> .....	250
The Nervous Regulation of Visceral Processes by <i>Detlev W. Bronk</i> .....	261
The Argument for Chemical Mediation of Nerve Im- pulses by <i>Walter B. Cannon</i> .....	276



**Part I. Progress in the Application of  
Physical Chemistry to Medicine**



# SOME ASPECTS OF THE COLLOID CHEMISTRY OF MEMBRANES

BY

Herbert Freundlich

*Distinguished Service Professor,  
University of Minnesota*

It is a truism to say that membranes are most important in biological processes. Nevertheless our knowledge of the properties and behavior of membranes is still rather scant. This may seem surprising if we call to mind the fair amount of work done on osmotic pressure in a state of equilibrium, on Donnan equilibria, etc. Actually these are the only phenomena treated even in good and comprehensive textbooks of physical chemistry. They represent, however, only one aspect of the effectiveness of membranes, a very limited aspect too, in my opinion, and one which is less important for biological processes than some others.

Membranes set up barriers. In a limiting case they may be strictly semipermeable, allowing the solvent to pass but preventing the solute from doing so. Many intermediate cases, however, are possible and known. If a membrane separates two solutions containing substances which would react with each other, it prevents the straightforward reaction that would go on in a homogeneous solution and may cause an involved series of processes depending on the special permeability of the membrane. This permeability may be very complex: all substances present may be able to pass the membrane, but with greatly different velocities; the membrane may be asymmetrical, allowing a substance to pass with ease from the one side, with difficulty from the other; or it may be permeable to cations from the one side, to anions from the other, etc. This complex behavior is fully realized by biologists when treating natural membranes. To a less degree, only occasionally, and rarely with

## HERBERT FREUNDLICH

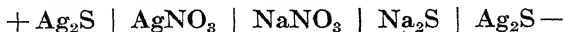
any thoroughness have physicochemists investigated these phenomena in artificial membranes, although some striking cases were observed as many as fifty to one hundred years ago by the first workers (1) in this field. Generally, as was just mentioned, physicochemists are interested in using semipermeable membranes only as means to a certain end, namely, for determining osmotic pressures in equilibrium. As soon as we investigate these processes in detail—their mechanism, their kinetics—we have to go into problems concerning the structure of membranes, their porosity, their electrical properties, etc. This means, since membranes are as a rule thin layers of a gel, that we have to treat them from the point of view of colloid science. My intention is to do so, but to give a brief survey of the complex membrane processes mentioned only in so far as they concern artificial membranes, without dealing with the orthodox equilibrium of osmotic pressure and with semipermeable membranes.

I shall begin with an example which probably does not allow any application to biological processes but which presents some essential facts in a clear, straightforward way and which shows how active a part artificial membranes may assume even in test-tube experiments. I am referring to the so-called Becquerel phenomenon, first investigated by Becquerel (2) in the 1860's and finally cleared up mainly by Bikerman (3) a few years ago. A test tube with a crack in it is filled with a  $\text{Na}_2\text{S}$  solution and dipped into a vessel containing a  $\text{AgNO}_3$  solution. When the two solutions meet in the crack, a membrane of  $\text{Ag}_2\text{S}$  is formed. Owing to the action of this membrane a beautiful druse of silver crystals is produced on the  $\text{AgNO}_3$  side of the test tube in the course of hours or days. On the  $\text{Na}_2\text{S}$  side  $\text{S}^{2-}$ -ion is oxidized to  $\text{S}_2^{2-}$ -ion; the solution turns yellow.

This phenomenon is due to the following series of causes: A strong membrane potential is set up between the two solutions and bridged over by two different paths of electric conductance; the  $\text{Ag}_2\text{S}$  is a metallic conductor; the pores of the membrane, which are filled with a  $\text{NaNO}_3$  solution, conduct elec-

## THE COLLOID CHEMISTRY OF MEMBRANES

trolytically. Hence we have at first the following galvanic cell acting in a closed circuit (see Figure 1):



and after some time, when silver has been electrolytically deposited on the  $\text{AgNO}_3$  side:



It is this electrolytic process that leads to the formation of the silver crystals. The two solutions remain separated owing to the

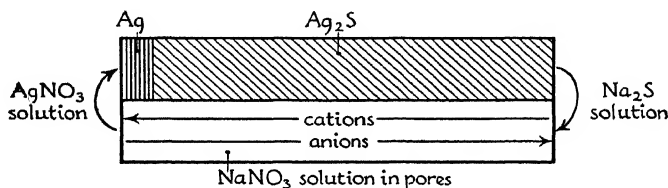


Figure 1. Diagram of Becquerel phenomenon. (Drawn in the Medical Art Shop, University of Minnesota.)

special permeability of the membrane: it is permeable to cations, impermeable to anions from the  $\text{Na}_2\text{S}$  side; permeable to anions, impermeable to cations from the  $\text{AgNO}_3$  side. Hence neither  $\text{Ag}^+$ - nor  $\text{S}^{=}$ -ion pass through. This permeability could be proved by direct chemical analysis after the addition of suitable, easily analyzed ions ( $\text{N}(\text{CH}_3)_4^+$ -,  $\text{HClO}_4^-$ -ion, etc.) to the one or the other side. This is a purely electrical effect, the potential gradient across the membrane, about 25 volt/cm., being large enough to prevent the cations from migrating against the current, and the anions in the opposite direction. The pores of the  $\text{Ag}_2\text{S}$  membrane are not particularly narrow; they are actually rather wide, allowing large neutral organic molecules to pass through and having no osmotic effect.

Some fairly general conclusions can be drawn from these results: If the two solutions had been mixed directly,  $\text{Ag}_2\text{S}$  would have been precipitated rapidly; the only change that would have gone on in course of time would have been an increase in size of the  $\text{Ag}_2\text{S}$  crystals. The membrane forces the difference in

## HERBERT FREUNDLICH

free chemical energy of the two solutions to look for adjustment on a quite different route, causing the formation of different substances in a special distribution in space. It initiates, as it were, an interlude which after a very long time would also lead to the formation of  $\text{Ag}_2\text{S}$  alone. The mechanism producing this interlude is attributable to the closed electric circuits generated on the membrane that separates the two solutions. These circuits can develop if the membrane has, so to speak, a checkered or mosaic structure, small blocks of different properties lying next to each other. In the membrane in question the parts consisting of  $\text{Ag}_2\text{S}$ , having metallic conductance, and the pores, conducting electrolytically, represent these different blocks.

It is perhaps interesting to add that Ostwald (4) tried in vain to give a consistent explanation of Becquerel's phenomenon; the reason for his failure was that he considered the membrane to be a thin, homogeneous layer of  $\text{Ag}_2\text{S}$  and that he took no account of the pores.

Membrane phenomena, essentially of this type, may vary in manifold ways. The Becquerel phenomenon is peculiar in that the membrane has in part a metallic conductance; the closed electric circuits therefore pass partly through the solid mass of the membrane, and are further able to cause electrolytic deposition. In most membranes we have no metallic conductivity; their solid parts are generally bad conductors of electricity. We may nevertheless have closed electric circuits, as was first emphasized by Sollner (5), namely, liquid closed circuits. In 1906 Dolezalek and F. Krüger (6) experimented as follows: A glass ring (see Figure 2) was filled with three aqueous solutions, a solution of  $\text{H}_2\text{SO}_4$  touching on the one side a solution of  $\text{LiCl}$ , on the other a solution of  $\text{Na-acetate}$ . This combination would lead one to expect that the diffusion potentials at the junctions of the solutions would result in a fairly large potential. The latter would cause an electric current to pass around the ring as long as there were still differences in concentration in the liquid. This was shown to be true by the deflection of a magnetic needle brought inside the ring. Such liquid closed electric



## THE COLLOID CHEMISTRY OF MEMBRANES

circuits can occur in a membrane separating two electrolyte solutions—in the simplest case, solutions of one electrolyte in different concentrations—provided the membrane is checked in a suitable manner. Actually it is sufficient that the membrane is not homogeneous as to pore diameter, but contains very fine pores next to wider ones (Sollner, 5).

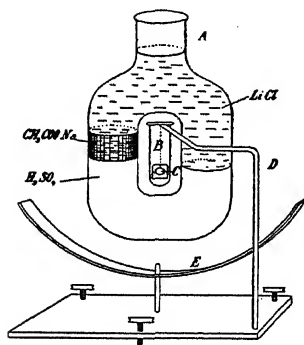


Figure 2. Diagram of experiment proving liquid closed electric circuit.

Here another property of membranes, first investigated by Michaelis (7), becomes important. Michaelis showed that, for instance, a collodion membrane markedly influenced the mobility of ions if it had been dried thoroughly and if the diameter of the pores was correspondingly small. The effect depends on the electrokinetic potential of the wall of the membrane pores. If it is negative, as in collodion, the anions are much more strongly retarded than the cations. The difference in the mobilities of the cations is also pronouncedly increased; the ratio of the mobilities of  $H^+$ -ion and  $Li^+$ -ion is about ninety times larger in a collodion membrane than in free solution. In positive membranes the cations are much more strongly retarded than the anions. Now the diffusion potential of an electrolyte in free solution and also across a membrane depends on the difference in the mobilities of its cation and anion. For  $KCl$ , for instance, the diffusion potential in free solution and in wide pores is practically zero, because the mobilities of the two ions are nearly

## HERBERT FREUNDLICH

equal. If, however, a collodion membrane with very fine pores separates two KCl solutions of different concentration, e. g., 0.01 and 0.001 respectively, a marked diffusion potential—up to 50 millivolts—will develop, because the rate of migration of the

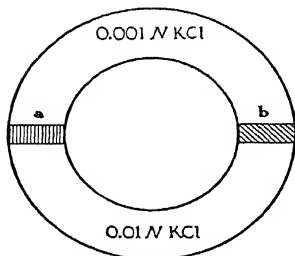


Figure 3. Diagram of liquid closed electric circuit produced by membranes. (Drawn in the Medical Art Shop, University of Minnesota.)

$K^+$ -ion is retarded much less than that of the  $Cl^-$ -ion. If we separate in Dolezalek's ring two KCl solutions of different concentration, on the one side by a membrane with fine pores (*a* in Figure 3), on the other side by one with wide pores (*b* in Figure 3), a liquid closed electric circuit is set up. And we have the same arrangement if a membrane with fine and wide pores next to each other separates the two KCl solutions.

These circuits cannot produce any electrolytic deposition on membranes having no metallic conductance, but they can cause another phenomenon—electrosmosis—if there is the possibility of transport of liquid owing to the circuits. The transport of liquid through the membrane produced by this electrosmosis is quite independent of the normal transport caused by the osmotic pressure. This phenomenon is known as abnormal osmosis. It cannot lead to a well-defined equilibrium as in normal osmosis, because the membrane is not allowed to be strictly semipermeable. The final state is an equal distribution of the solute between the two compartments. Abnormal osmosis, however, may have strong kinetic effects, when conditions are far from the final state and in dilute electrolyte solutions, where it may outweigh normal osmosis. The transport of liquid by abnormal osmosis may be in the same direction as that by normal osmosis, that is, from the dilute solution to the concentrated one; this is positive abnormal osmosis. Or the transport may be in the opposite direction, from the concentrated solution to the more dilute one; this is negative abnormal osmosis. It has been

## THE COLLOID CHEMISTRY OF MEMBRANES

shown by Sollner (5) that all experimental results (8) obtained so far concerning positive and negative abnormal osmosis, when using nonswelling membranes (collodion, baked clay, Mg-silicate, etc.) — results such as the influence of the nature of the ions, the influence of the electrokinetic potential of the membrane, etc. — can be explained on this basis, that is, as being caused by the electrosmotic movement of the liquid produced by liquid closed electric circuits. It has further been shown by Sollner and Grollman (5) that actually the membrane potential of a Mg-silicate membrane which had pores of a certain average size could produce negative osmosis, that is, it could move a LiCl solution from the concentrated to the dilute compartment through a second Mg-silicate membrane having pores of a different average size.

It is a fairly simple type of inhomogeneity, of mosaic structure, which we have to take into account in the phenomena just discussed. Under biological conditions the membrane effects may be caused in most cases by much more involved mosaic structures. Actually the conception of a mosaic structure, when first introduced by Nathansohn (9) in 1904, referred to a more complex membrane. He explained the fact that while natural membranes frequently behaved as lipoid membranes — i.e., they were mainly permeable to fat-soluble substances — they generally also had a certain permeability for water-soluble substances like salts, by assuming that the membranes consisted of small blocks of, let us say, lipoids and proteins next to each other.

As was mentioned above, the final state is an equal distribution of the electrolyte in both compartments. The transport of liquid due to abnormal osmosis goes on only as long as the membrane potential and the closed circuits are active. This transport of liquid is not the only change occurring; Bethe and Toropoff (10) have shown that when a salt solution is moved through a membrane by electrosmosis, marked changes in acidity and alkalinity are observed close to the membrane, owing to the particular mobilities of  $H^+$ - and  $OH^-$ -ion and the fact, mentioned

## HERBERT FREUNDLICH

above, that they are altered comparatively little by membranes with fine pores. There is no reason why the electrosmosis due to liquid closed circuits should not have the same effect. Hence we would have not only a temporary change in the distribution of the mass of liquid in space but also one in the distribution of the different ions in solution: the membrane causes again a complicated interlude, again with a marked geometrical factor, a special distribution of the substances in space, before the final state is reached.

With membranes that are able to swell, abnormal and particularly negative osmosis may be produced by another mechanism, namely, by differences in the rate of swelling, as was shown by Flusin (11). If a rubber membrane separates two liquids in which it swells with different velocities, e. g., benzene and alcohol, the liquid is transported osmotically through the membrane from that side of it which swells more quickly, in our example from the benzene side. In these experiments the rubber membrane must be prevented from bulging out by pressing it between two nettings of thin wire. The same situation holds with a membrane like pig's bladder between water and an aqueous solution: if the aqueous solution, e. g., one of tartaric acid, causes the membrane to swell more quickly than in water, the transport of liquid is from the acid solution to the pure water, that is, we have a negative osmosis. This behavior of pig's bladder between water and an aqueous solution of an organic acid was observed by Dutrochet (1) but the phenomenon has not yet been thoroughly investigated. In particular only Bartell (12) seems to have discussed the question of how behavior of membranes that are able to swell may be correlated to that of nonswelling membranes, as treated above.

Finally there may be briefly discussed another group of phenomena, also able to cause a special distribution of substances across a membrane differing from that produced by normal osmosis. These phenomena are not strictly dependent upon the presence of membranes, but are strongly enhanced by them.

We may start from the simple Donnan equilibrium: A solu-

## THE COLLOID CHEMISTRY OF MEMBRANES

tion of a colloidal electrolyte, i. e., of an electrolyte having one very large ion, like congo red, on the one side of a membrane that is impermeable to the colloidal ion, and a true solution, like one of NaCl, on the other side. In equilibrium the NaCl is not distributed equally on both sides; the presence of the colloidal ion allows only a comparatively small amount of NaCl to pass to the other side. This Donnan effect is an example of a very general phenomenon, as was first emphasized by Hartley (13). If a liquid solution is in a field of force that is not uniform, the dissolved substance does not remain in uniform distribution throughout the liquid but becomes distributed nonuniformly, too. We need not be dealing with a state of equilibrium; the state may also be a stationary or a temporary one, kept up only as long as the gradient of force is sufficiently steep. The simplest phenomenon of this kind is perhaps the Ludwig-Soret effect (14): If a column of a solution having at the outset a uniform concentration is kept warm on the one end and cold on the other, the concentration of solute on the warm end becomes different from the concentration on the cold end. It was believed that this phenomenon could be explained simply by van't Hoff's conception of solutions. Assuming that the osmotic pressure would have to be equal throughout the column, one might expect that the increase of osmotic pressure caused by the higher temperature would be compensated by a corresponding decrease in concentration. This theory, however, did not agree with the experiments. Not only the molecules of the solute migrate under the influence of the temperature gradient, but also the molecules of the solvent. The process depends on their mutual affinity, and it is therefore correlated in a still unknown way to the solubility of the solute and the temperature coefficient of its solubility.

At constant temperature the presence of a second substance which is not uniformly distributed, but which has a concentration gradient, is sufficient to produce a nonuniform field of force. The concentration gradient of the second solute causes an unequal distribution of the first solute, which originally was uni-

## HERBERT FREUNDLICH

formly distributed. This phenomenon is known as anomalous diffusion (15) or *diffusion rétrograde*. One example may be given (16): If  $K_2SO_4$  is allowed to diffuse in a uniform solution of quinone, the distribution of the quinone becomes nonuniform where the concentration gradient of the salt is steep; the quinone concentration decreases in the sulfate solution and increases in the purely aqueous solution.  $K_2SO_4$  is a substance

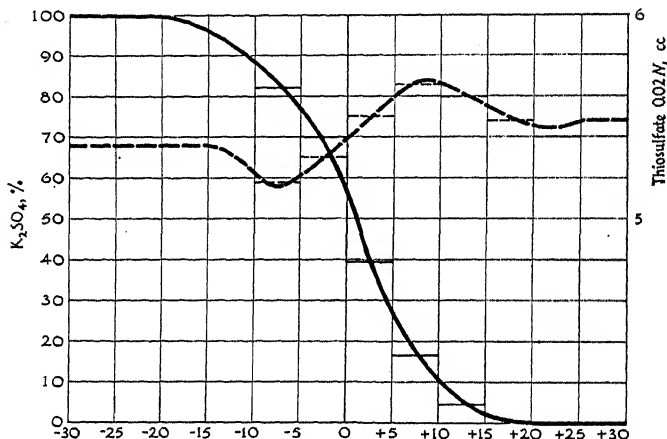


Figure 4. Anomalous diffusion of quinone caused by a concentration gradient of potassium sulfate. (Drawn in the Medical Art Shop, University of Minnesota.)

which causes a decrease in the solubility of quinone;  $KNO_3$ , which causes an increase in the solubility of quinone, alters its distribution in the opposite direction.

Figure 4 is a graphic representation of this behavior. The abscissa gives the twelve layers of the diffusion column, the majority of which were analyzed separately; point 0 is the junction of the two solutions, the one containing at the outset  $K_2SO_4$  + quinone, the other quinone alone; 25-30 is the top layer. The solid curve refers to the distribution after ten hours time of the diffusing  $K_2SO_4$  and to the figures on the ordinate to the left

## THE COLLOID CHEMISTRY OF MEMBRANES

expressed in percentages of the original  $K_2SO_4$  concentration, which was 0.827*N*. The broken curve refers to the distribution of quinone and to the figures on the ordinate to the right, which are in cubic centimeters of a 0.02 thiosulfate solution needed for titrating 2 cc. of each layer in question. The concentration of the quinone in the two solutions at the outset was not exactly the same; the titration values were 5.36 for the solution containing  $K_2SO_4$ , 5.50 for the purely aqueous solution. The absolute concentration of the quinone was about 27 millimoles per liter.

The Donnan effect is a special case of anomalous diffusion where a semipermeable membrane allows an equilibrium to be reached. Even if we are not dealing with a strictly semipermeable membrane, membranes as a rule will make phenomena of this kind more pronounced, retarding diffusion and thus keeping up a steep concentration gradient for a longer time. Not much work has been done along these lines on the more general and nonisothermal effects. Lippmann (17) and Aubert (18) have investigated thermosmosis, i. e., the transport of liquid occurring when a membrane separates two compartments containing the same pure liquid at different temperatures. Effects may be in either one or the other direction, the influence of the nature of the membrane and of the presence of foreign substances in small amounts in the membrane being remarkable. It seems probable, according to Sollner (19), that fairly small differences in temperature may also cause appreciable changes in the distribution of substances across membranes when one is dealing with solutions rather than pure liquids.

If one is dealing with isothermal conditions the following fact deserves to be emphasized, particularly from a biological point of view: Provided one of the substances present has a concentration gradient, this will produce or favor an asymmetrical distribution of other substances, which from the outset may have been more uniformly distributed. It is perhaps doubtful whether effects of this kind are large enough to explain the strong antiosmotic distribution of substances frequently found

## HERBERT FREUNDLICH

in living organisms. Presumably some specialized mechanisms, "two-dimensional organs" according to Krogh (20), will have to be taken into account. Nevertheless the possible influence of anomalous diffusion ought not to be neglected.

It is characteristic of life that it is a stationary state of chemical and physicochemical processes at a distance from equilibrium. Life is also a very long and, if we consider propagation, a practically eternal interlude. Membranes favor interludes strongly interrupting the straightforward course of reactions. It is another characteristic feature of biological processes that they do not occur, as a rule, in homogeneous solutions, but that chemical reactions are of necessity closely correlated to the existence of a special structure in space. Membranes favor a particular distribution of substances in space; they underline, so to speak, a geometrical factor. These facts form perhaps the essential part of the influence which membranes bring to bear upon biological processes.

### *References*

1. Dutochet, R. J. H. 1835. *Ann. de chim. et de phys.* 60:337; Graham, Thomas. 1854. *Phil. Tr. Roy. Soc. London* 144:177.
2. Becquerel, A. C. 1867-77. *Compt. rend. de acad. de sci.* Vols. 64, 65 (1867); Vols. 66, 67 (1868); Vol. 71 (1870); Vol. 74 (1872); Vol. 76 (1873); Vols. 78, 79 (1874); Vol. 80 (1875); Vol. 82 (1876); Vols. 84, 85 (1877).
3. Bikerman, J. J. 1931. *Zeitschr. f. physik. Chem. A* 153:451; 1933. *Kolloid-Zeitschr.* 62:26; cf. also Freundlich, H., and Sollner, K. 1928. *Zeitschr. f. physik. Chem. A* 139:349; 1931. *Zeitschr. f. physik. Chem. A* 152:313.
4. Ostwald, Wilhelm. 1890. *Zeitschr. f. physik. Chem.* 6:71.
5. Sollner, K., and Grollman, A. 1932. *Tr. Electrochem. Soc.* 61:477, 487; Sollner, K. 1930. *Zeitschr. f. Elektrochem. u. angew. physik. Chem.* 36:36, 234; Sollner, K., and Grollman, A. 1932. *Zeitschr. f. Elektrochem. u. angew. physik. Chem.* 38:274.
6. Dolezalek, F., and Krüger, F. 1906. *Zeitschr. f. Elektrochem. u. angew. physik. Chem.* 12:669.
7. Michaelis, L., and collaborators. 1925. *Jour. Gen. Physiol.* 8:33; 1927. *Jour. Gen. Physiol.* 10:575, 671, 685; 1928. *Jour. Gen. Physiol.* 11:147; 1929. *Jour. Gen. Physiol.* 12:55, 221, 473, 487; 1928. *Colloid Sympos. Monogr.* 5:135; 1925. *Biochem. Zeitschr.* Vols. 158, 159, 161, 162, 164; 1926. *Biochem. Zeitschr.* Vol. 170.



## THE COLLOID CHEMISTRY OF MEMBRANES

8. Bartell, F.E. 1914. Jour. Am. Chem. Soc. 36:446; 1920. Proc. Nat. Acad. Sci. 6:308; 1923. Colloid Sympos. Monogr. 1:120; Bartell, F. E., and collaborators. 1916. Jour. Am. Chem. Soc. 38:1029, 1036; 1920. Jour. Phys. Chem. 24:444, 593; 1923. Jour. Phys. Chem. 27:101, 252, 346.
9. Nathansohn, A. 1904. Jahrb. d. Wissensch. Botanik 39:607.
10. Bethe, A., and Toropoff, T. 1914. Zeitschr. f. physik. Chem. 88:686; 1915. Zeitschr. f. physik. Chem. 89:597.
11. Flusin, G. 1908. Ann. de chim. et de phys. (8) 13:480.
12. Bartell, F. E., and Sims, L. B. 1922. Jour. Am. Chem. Soc. 44:289.
13. Hartley, G. S. 1931. Tr. Faraday Soc. 27:10.
14. Ludwig, C. 1856. Ber. f. Wien. Akad. d. Wissensch. 20:539; Soret, C. 1879. Arch. Genève 3:48; 1880. Arch. Genève 4:209; 1881. Ann. de chim. et de phys. (5) 22:293; cf. also Hartley, G. S. 1931. Tr. Faraday Soc. 27:1.
15. Thovet, J. 1914. Ann. de phys. (9) 2:369; cf. also Hartley, G. S. 1931. Tr. Faraday Soc. 27:10.
16. Freundlich, H., and Krüger, D. 1939. Jour. Phys. Chem. 43:981.
17. Lippmann, G. 1907. Compt. rend. de acad. de sci. 145:104.
18. Aubert, M. 1912. Ann. de chim. et de phys. (8) 26:145, 551.
19. Private communication. Cf. also Ernst, E. 1939. Kolloid-Zeitschr. 87:276.
20. Krogh, A. 1937. Tr. Faraday Soc. 33:912.

# THE PERFORMANCE OF OSMOTIC WORK IN LIVING SYSTEMS

BY

Maurice B. Visscher

*Professor of Physiology, University of Minnesota*

LIVING systems may be described from a physicochemical point of view as systems maintained in a steady state at positions far from true equilibrium. The physiologist Claude Bernard recognized this fact when he stated the law of constancy of the internal environment as a condition of life. Cannon's concept of homeostasis puts the same idea forward as a unifying principle in physiology.

In chemical systems such as one finds in a living organism the maintenance of a steady state at some distance from the position of true reversible equilibrium can be accomplished only by the continuous expenditure of energy. There are many aspects of the problem of steady states important to biology but I shall confine myself strictly to observations concerning the osmotic activity of living systems.

In order to point out that osmotic work is of more than academic interest I shall mention briefly some applications. From a medical point of view the most conspicuous example is the secretion of urine. The kidney has the largest responsibility for maintaining osmotic constancy in mammalian forms. In lower vertebrate forms, as Keys has shown, this function is handled mainly by certain specialized structures in the gills. In mammals the kidney is capable of concentrating substances a hundredfold in urine as compared with blood, and it can secrete urine with an osmotic pressure as much as twenty atmospheres greater than that of plasma. Moreover, it is the concentration of individual constituents rather than the over-all osmotic pressure which is of the highest importance.

## OSMOTIC WORK IN LIVING SYSTEMS

The selectivity with which the kidney does work must be emphasized. From a practical point of view the clinician is mainly interested in the pathological physiology of the kidney, and is therefore concerned with those conditions in which the kidney fails to do selective osmotic work in a normal manner. Analyzed broadly in physicochemical terms a renal insufficiency constitutes such a failure. Specifically this defect may result from the anatomical destruction of certain units, or it may be the result of chemical changes in the cells of the kidney tubules which leave no visible microscopic sign.

It might seem to be of no great consequence to know that when a large proportion of glomerular and tubular units in a kidney are destroyed, as in glomerulo-nephritis, the essential functional pathology is in the osmotic work mechanism. It might seem that since the damage is obviously structural in a microscopic or even a gross anatomic sense the resultant functional defect would be so obvious as not to require study. Actually, however, this is not true. The kidney like other organs has a large margin of safety in its capacity to do work, and a large fraction, even a major portion, of its mass can be destroyed or removed without impairment of function except under extreme conditions. It is the business of the physician to assess the degree of functional damage by studying the limits of capacity to do osmotic work and to govern the treatment of his patient accordingly. Every renal function test represents fundamentally a measure of the ability of the renal mechanism to perform osmotic work. The dilution-concentration tests, the various clearance measurements, and the dye excretion methods all involve estimations of the ability to do certain types of osmotic work.

In other disease states the ability of the kidney to do osmotic work is impaired without obvious renal histopathology, as for example in adrenal cortical insufficiency, where the concentrating power of the kidney tubules for numerous urinary constituents is much reduced. The defect of the kidney tubules in this disease may be shared by other cells in the body, for all active

## MAURICE B. VISSCHER

cells probably do work to maintain the normal distribution ratios between intra- and extracellular fluids. It is not impossible that in adrenal insufficiency the body cells generally are less able to do the osmotic work by which they maintain their individual chemical characteristics.

For several years my research associates and I have been engaged in investigating the circumstances surrounding the performance of osmotic work in various living systems. We have studied a number of gland secretory processes and the maintenance of concentration differences between cells and tissue fluids, but we have devoted our main attention to absorption from the intestine. Intestinal absorption has in the past been considered generally to involve mainly passive processes; but it is now obvious that although passive processes also take place, a very important part of intestinal absorption occurs against concentration gradients, and that osmotic work is done in such movement.

The lower ileum can be made to do osmotic work which is qualitatively and quantitatively somewhat similar to that done by the renal tubules. The intestine offers important advantages for investigation over the kidney tubule because it is a macroscopic rather than a microscopic organ and therefore lends itself better to investigation. Although the mechanism by which the intestine performs osmotic work is important in its own right the main objective of our studies has been to discover whatever empirical regularities may be found, with the expectation that the fundamental mechanisms are apt to be similar in comparable situations elsewhere in living systems.

We have studied in detail concentration and volume changes in active absorption. Complete balances have been calculated in order to eliminate the possibility of simple ion exchanges. The influences of numerous factors, such as (H), anesthetics and poisons, species of animal, and level of the small intestine, have all been studied. The movement of water has been investigated by the use of  $D_2O$ , by the study of total osmotic activity, as well as by volume measurements. We have determined the

## OSMOTIC WORK IN LIVING SYSTEMS

hydrostatic pressures in the gut during absorption to rule out simple filtration processes. The absorption of the animal's blood serum introduced into the intestinal loop has been studied, in order to set up conditions in which all concentrations are identical on the blood and intestinal lumen sides of the gut wall. The absorption of a wide variety of substances moving both against and with diffusion gradients has been studied.

From the results of this work certain general rules can be derived. As a generalization it can be stated that when mixtures of nontoxic uni-univalent and uni-polyvalent or poly-univalent salt solutions are placed in the small intestine, the uni-univalent salt will be rapidly absorbed against concentration gradients. As an example one can take the case of an equiosmotic isotonic mixture of  $\text{NaCl}$  and  $\text{Na}_2\text{SO}_4$ . When such a solution is placed in a loop of ileum one notes (see Figure 1) that the  $\bar{\text{Cl}}$  is almost completely removed. The  $\bar{\text{Cl}}$  is moved against concentration gradients as great as 200:1, and simple calculation in experiments where all important constituents have been measured

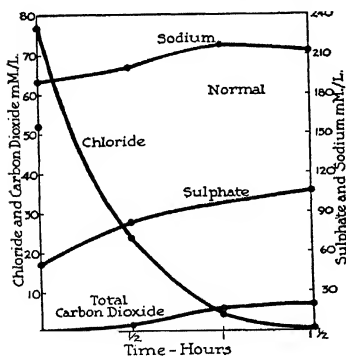


Figure 1 (left). Concentrations of chloride, sodium, sulfate, and total carbon dioxide in fluid originally approximately isotonic with blood placed in a loop of lower ileum in a dog anesthetized with amytal. (From Ingraham and Visscher, 1936. Drawn in the Medical Art Shop, University of Minnesota.)

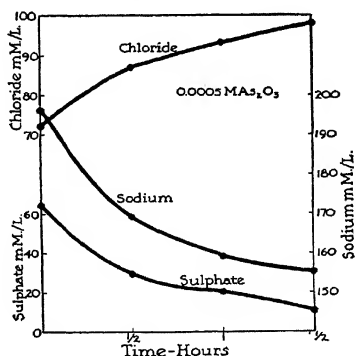


Figure 2 (right). A comparable experiment on a loop of gut adjacent to that used in the experiment shown in Figure 1. The only alteration in the procedure was the addition of 0.0005 M  $\text{As}_2\text{O}_3$  to the loop fluid. (Drawn in the Medical Art Shop, University of Minnesota.)

## MAURICE B. VISSCHER

shows that in approaching the steady state the system has moved away from the conditions of true equilibrium. Figure 2 shows the results of a similar experiment, in which  $\overset{+}{\text{Na}}$  and  $\overset{-}{\text{SO}}_4$  were also measured in the fluids from two adjacent loops of lower ileum, to one of which 0.0005M  $\text{Na}_2\text{HASO}_3$  had been added. Similar effects are obtained by the use of  $\text{HCN}$ ,  $\text{H}_2\text{S}$ , and  $\text{HgCl}_2$ .

Univalent salt impoverishment has been observed not only with  $\overset{-}{\text{SO}}_4$  as the polyvalent anion but also with  $\overset{-}{\text{HPO}}_4$ , Citrate and Ferrocyanide; and the corresponding effect with the cations  $\overset{+}{\text{Mg}}$ ,  $\overset{+}{\text{Mn}}$ , trivalent Cobaltamine, and  $\overset{+}{\text{Ca}}$  — the last only in fairly acid solutions.

A series of studies has been conducted to investigate the relation between water movement and salt transport during what may be called *active absorption* from the intestine. Conventional reasoning has assumed that the volume change in absorption may be taken as a reliable measure of total water movement, neglecting the possibility that there may be significant movement into the intestine from the blood, perhaps through the intestinal glands. There could be a volume decrease with fluid both entering and leaving if the rate of exit exceeded the rate of entrance of water. In an attempt to determine whether there is movement in two directions we have made use of heavy water, which, added either to the blood or the intestinal fluid, permits a study of one-way movement. By placing salt solutions containing  $\text{D}_2\text{O}$  in the intestine it is possible to compare the rates of clearance of, for example,  $\overset{-}{\text{Cl}}$  and  $\text{D}_2\text{O}$  simultaneously. The clearance factor is defined as the volume of fluid in the gut under standard experimental conditions, which is completely cleared per unit of time of any component. The interesting fact comes out of these observations (see Figure 3) that the clearances of  $\overset{-}{\text{Cl}}$  and  $\text{D}_2\text{O}$  are found to be very nearly equal when the  $\overset{-}{\text{Cl}}$  clearance is high. This approach to equality at high rates of active absorption might be simply a coincidence, but on the other hand it more likely implies a

## OSMOTIC WORK IN LIVING SYSTEMS

fundamental connection between the mechanisms of movement of the two substances. The results would seem to favor the latter view, since the NaCl moved against and the D<sub>2</sub>O with concentration gradients. The results might therefore be taken to mean that a much larger volume of water moved from gut to blood than would be indicated by the observed volume changes. Such a movement could occur only if a relatively large volume of fluid were entering the intestine during the process.

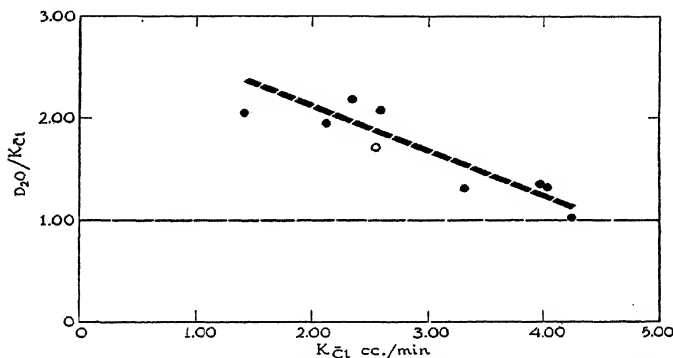


Figure 3. The ratio of the rates of clearance of D<sub>2</sub>O and Cl plotted against the absolute rate of Cl clearance. Each point represents a specific experiment in which an isotonic solution containing NaCl and Na<sub>2</sub>SO<sub>4</sub> in water containing approximately 4 per cent D<sub>2</sub>O was placed in a loop of lower ileum and absorption studied for 10 to 17 minutes. (For further explanation see original paper, Peters and Visscher, 1939. Drawn in the Medical Art Shop, University of Minnesota.)

By the use of the Baldes-Hill thermocouple method Dr. R. R. Roepke and I have followed the vapor tension of the fluids involved as measures of their total osmotic activity. It has been found as a general rule that when active absorption of uni-univalent salt is occurring from isotonic mixtures of the type referred to previously, the process is always associated with a decrease in osmotic activity of the gut fluid. This decrease may be as great as a half atmosphere osmotic pressure. Figure 4 shows the results of observations upon two loops of gut into which fluids were placed whose osmotic activities were equal to

## MAURICE B. VISSCHER

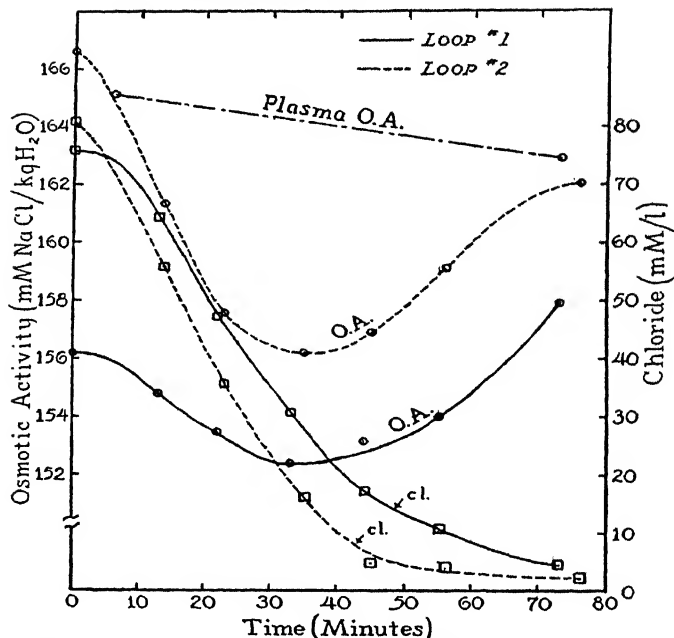


Figure 4. The course of changes in chloride concentration and the osmotic activity expressed as equivalents of NaCl concentration in two adjacent ileal loops in the plasma in a dog anesthetized with nembutal. The loops were filled with an approximately isotonic solution containing equiosmotic quantities of NaCl and Na<sub>2</sub>SO<sub>4</sub>. (From Roepke and Visscher, 1939. Drawn in the Medical Art Shop, University of Minnesota.)

or a little less than that of the plasma. The only possible conclusion is that the net movement of osmotically active solutes into the blood is correspondingly greater than the net movement of water. It is however impossible to decide on the basis of evidence of the sort just presented whether a hypertonic solution is absorbed or a hypotonic solution moved (secreted) into the gut. Either process or perhaps both occurring at once would yield the observed result.

Considering the total volume change it is possible to throw



## OSMOTIC WORK IN LIVING SYSTEMS

some additional light on the problem. Ordinarily there is a volume decrease associated with absorption of the type discussed. However in several instances we find active absorption with no change or even an increase in volume. In one such instance, shown in Figure 5, considerable water entered the intestine from the blood, while the osmotic activity of the gut fluid was always lower than that of the blood. This observation is cru-

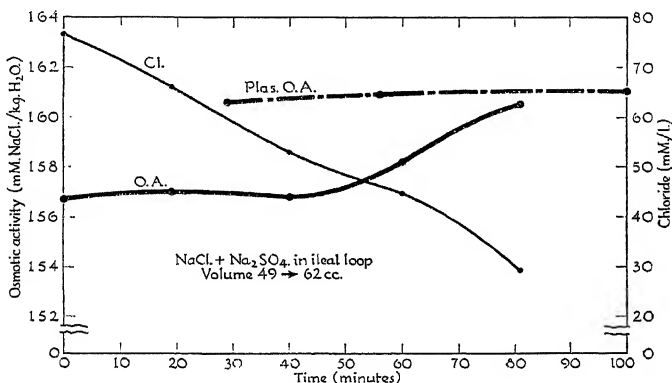


Figure 5. The course of chloride concentration and osmotic activity changes in fluid containing equiosmotic proportions of NaCl and  $\text{Na}_2\text{SO}_4$ , originally slightly less than isotonic, placed in a lower ileal loop. An increase in loop fluid volume is noted despite an osmotic gradient in the direction which should result in the passive movement of water from the loop to the blood. (Drawn in the Medical Art Shop, University of Minnesota.)

cial to our reasoning. There can be no question that the intestinal epithelium is capable of moving water against osmotic gradients, although it does not exhibit this property under ordinary circumstances—perhaps because the effect is usually masked by other actions.

As a final point concerning intestinal absorption mention may be made of the absorption of the animal's own serum placed in the gut. In this case there is obviously no difference in chemical composition or physicochemical properties of the fluid on the lumen and the blood sides of the intestine. Yet as Carl Voit first showed in 1869, absorption occurs. Now it is

## MAURICE B. VISSCHER

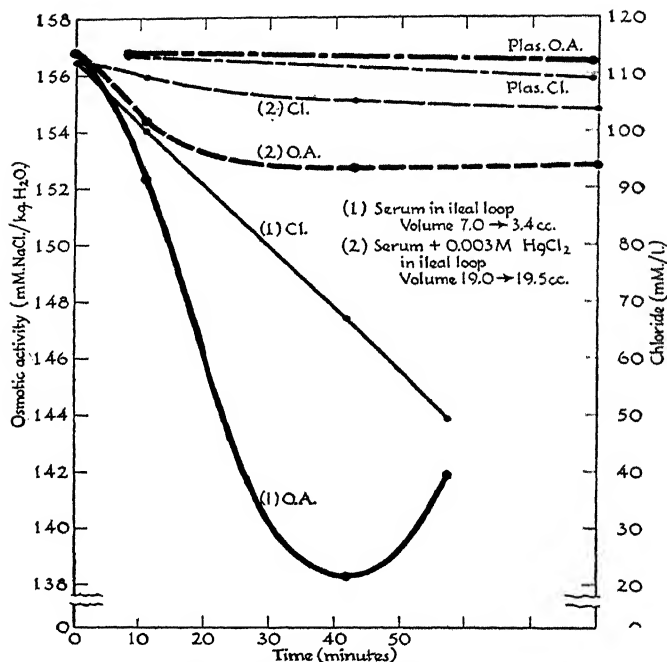


Figure 6. Observations upon chloride concentration and osmotic activity during absorption of serum from adjacent ileal loops, one of which was poisoned with 0.003 M HgCl<sub>2</sub>. (Drawn in the Medical Art Shop, University of Minnesota.)

seen (Figure 6) that during this absorption of serum its osmotic activity decreases. In the second loop poisoned with HgCl<sub>2</sub> this effect was nearly abolished. Again the net over-all movement into the blood stream of solute is relatively greater than that of solvent. Further it is to be noted that as time proceeds the analytical concentration of chloride decreases. These observations are important particularly because they completely exclude the possibility that passive diffusion processes are responsible for the osmotic activity changes.

To conclude the presentation of observational data I should

## OSMOTIC WORK IN LIVING SYSTEMS

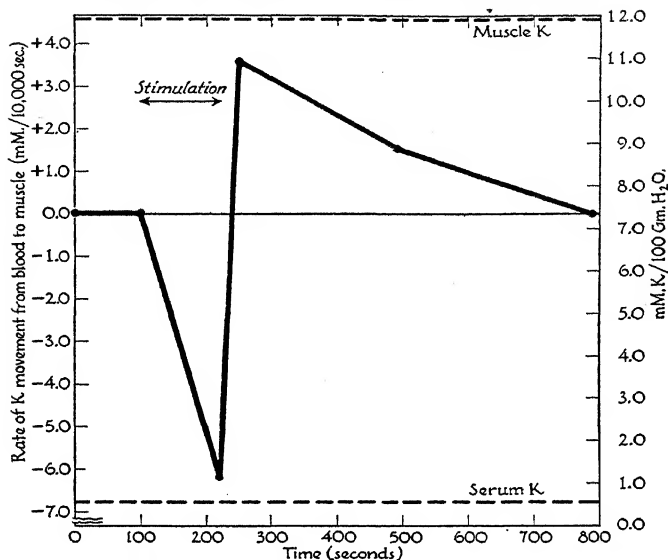


Figure 7. A graph showing the rate of movement of potassium between blood and muscle in the isolated dog's gastrocnemius perfused with a heart-lung preparation. The measured muscle and serum potassium levels in this experiment are shown. (This experiment is from the unpublished work of Earl H. Wood. Drawn in the Medical Art Shop, University of Minnesota.)

like to describe the results of an experiment by Earl H. Wood on an entirely different organ. I have referred before to the fact that every active cell in the animal body maintains a chemical composition different from that of the tissue fluid medium in which it lives, and that at least in the case of some of these cells the constituents move in and out. In Figure 7 appear the results of an experiment in which the isolated gastrocnemius of the dog was perfused with a heart-lung preparation. By measuring blood flow and the chemical composition of the arterial and venous blood, and, at the end of the experiment, that of the muscle, it is possible to follow the rate of movement of, in this case, potassium, in relation to concentration gradients. It will be seen that during rest potassium neither leaves nor enters

## MAURICE B. VISSCHER

the muscle. During and shortly after contraction, however, there is a large loss of potassium, as was first shown by Fenn. A few seconds after contraction ceases the potassium again moves into the muscle. Wood's experiments show for the first time and in a clear way how effectively the muscle cell reverses the direction of movement. For the purpose of this discussion I wish to focus attention upon the gradient against which potassium moves in recovery, as indicated by the difference between the plasma and the muscle concentrations. The gradient is actually steeper than these figures would indicate, because the low potassium content of the tissue spaces in the muscle is not taken into account. One might admit that a large share of the potassium in muscle is undissociated, and there would still be a large concentration gradient.

It might be suggested that the movement of  $K^+$  against twenty-fold gradients may be due simply to an ion exchange mechanism, in which an approach to thermodynamic equilibrium occurs. However, such a suggestion that no osmotic work is done ignores the most fundamental point that the setting up and maintenance of the necessary concentration of exchangeable ions involves in itself an expenditure of energy. This energy finally appears as osmotic work.

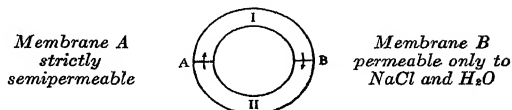
This discussion is necessarily brief and it is therefore impossible to enter into an extended analysis of how living membranes may do various types of osmotic work. Some of the more likely physicochemical processes have already been discussed by Professor Freundlich. He has referred primarily to the driving forces, the sources of energy, which might be put to work. As he has said, too little is known about the properties of the membranes involved in the performance of osmotic work to permit one to say how such work is accomplished in any living cell.

I should like to refer briefly, however, to a simple type of mechanism of a sort we know to occur in some situations in animal forms, which would be capable, if it exists in the situations in question, of performing the osmotic work done by the

## OSMOTIC WORK IN LIVING SYSTEMS

intestine. We have pointed out the obvious fact that if a mechanism exists in which fluid moves in a circuit through two membranes of differing permeability there can be a separation of constituents (see Figure 8). If membrane B, for example, is permeable to uni-univalent salts and impermeable to polyvalent ions, and membrane A is impermeable to any salt; and if com-

*In initial state compartment I contains  $\text{NaCl}$  and  $\text{Na}_2\text{SO}_4$ .  
Compartment II contains  $\text{NaCl}$ .*



*In final state all  $\text{NaCl}$  is in II and  $\text{Na}_2\text{SO}_4$  remains in I.  
The volume of II increases while that of I decreases.*

*Figure 8. A diagram illustrating the mechanism by which a separation of materials comparable to that observed in the case of intestinal absorption might occur as a result of fluid movement through a circuit involving membranes of differing permeability. (Drawn in the Medical Art Shop, University of Minnesota.)*

partment I originally contains a mixture of univalent and polyvalent ion salts, whereas II contains originally a univalent ion salt; then if the fluid is moved in the direction of the arrows all the uni-univalent salt would move to compartment II, and the fluid in compartment I would finally become a pure solution of polyvalent ion salt. This model can account satisfactorily for the course of events in intestinal absorption of mixed salt solutions described earlier. The relative volumes of compartments I and II are not fixed so that the volume of I decreases while II increases, allowing the total osmotic activities in each compartment to remain constant.

Of course other systems can be conceived which would also give the same result, but none with which I am acquainted is so simple in principle. Furthermore we know that mechanisms somewhat similar to the one postulated do occur in living systems. For example, protozoans such as *Amoeba proteus* and *Paramecium caudatum* rhythmically discharge from their con-

## MAURICE B. VISSCHER

tractile vacuoles quantities of fluid equal to as much as their cell volume in one hour. This process occurs continuously throughout the life of the organisms. As they do not lose volume it is obvious that an equal quantity of fluid is entering the organisms through other parts of the cell surface. Krogh (1939) concludes that nearly pure water must be excreted by the vacuole. The work that must be done by these organisms can readily be calculated, and it is found to be very large. The ability to perform such work has been found to be abolished by cyanide. According to Müller and others this process serves an osmoregulatory function.

It is not germane to my argument to insist upon the mechanism by which the contractile vacuole operates. I am content to rest on the simple observation that quantities of fluid equal to the volume of the entire organism pass from the external medium into the cell and out again through the vacuole as rapidly as once per hour. Here is obviously a fluid circuit in a very simple system. If osmotic work is done in the process the situation approximates the one under consideration in mammals even more closely, but that point is not crucial.

I should like to refer to another system in which osmotic work is known to be done and a kind of fluid circuit is known to exist, although in its simplest form it is not known to do all of the osmotic work. I refer to the vertebrate kidney itself. We know, on the basis of the work of numerous investigators, that a large volume of fluid is filtered through the glomerular capsules, a quantity about equal to the entire blood volume every hour. Ordinarily 99 per cent of this fluid returns to the blood in the course of its passage along the renal tubules. The high concentration of such substances as urea and creatinine in the urine is generally conceived to be due to the high degree of impermeability of the tubular epithelium to these substances which are therefore held back while the water moves to the blood. The reason for such impermeability and the mechanism by which the water is returned to the blood are unsolved problems, but the fact of a fluid circulation equal in volume to

## OSMOTIC WORK IN LIVING SYSTEMS

about a hundred times the eventual urine volume is not open to serious question.

It would not be strange if one were to find this mechanism, seen in the most primitive animal forms and in the most highly specialized organ found in nature for the performance of selective osmotic work, employed generally in accomplishing such work.

Future investigation in these fields must undoubtedly concern itself with the metabolic reactions which supply energy, perhaps in the form of concentration gradients and perhaps in other forms, which can serve as the driving forces for the performance of osmotic work.

### *References*

- Bernard, Claude. 1878. *Les Phenomenes de la vie*. Paris.
- Burns, H. S., and Visscher, M. B. Sept. 1934. The influence of various anions of the lyotropic series upon the sodium and chloride content of fluid in the intestine. *Am. Jour. Physiol.* 110:490-98.
- Cannon, Walter B. 1932. *The Wisdom of the Body*. New York.
- Ingraham, R. C., Peters, H. C., and Visscher, M. B. Oct. 1937. On the movement of materials across living membranes against concentration gradients. *Jour. Phys. Chem.* 42:141-50.
- Ingraham, R. C., and Visscher, M. B. 1937. Effects of a divalent cation on sodium removal from intestinal loops. *Proc. Soc. Exper. Biol. Med.* 36: 201-02.
- Sept. 1937. Further studies on intestinal absorption with the performance of osmotic work. *Am. Jour. Physiol.* 121:771-85.
- Oct. 1935. The influence of various poisons on the movement of chloride against concentration gradients from intestine to plasma. *Am. Jour. Physiol.* 114:681-87.
- Oct. 1935. The production of chloride-free solutions by the action of the intestinal epithelium. *Am. Jour. Physiol.* 114:676-80.
- Krogh, August. 1939. *Osmotic Regulation in Aquatic Animals*. Cambridge.
- Peters, H. C., and Visscher, M. B. Feb. 1939. On the mechanism of active absorption from the intestine. *Jour. Cell. Compar. Physiol.* 13:51-66.
- Roepke, R. R., and Visscher, M. B. 1939. Osmotic relationships between blood plasma and intestinal fluid during absorption. *Proc. Soc. Exper. Biol. Med.* 41:500-03.

# SOME REACTIONS BY WHICH SOLUTES MAY BE DIFFERENTIALLY CONCENTRATED BY THE KIDNEY

BY

John P. Peters

*Professor of Internal Medicine,  
Yale University*

WHEN the invitation was first given to me to address you, I was asked to discuss the work of the kidney. To have attempted this with only a scanty knowledge of the basic natural sciences and not even a command of the language of thermodynamics would have been presumptuous in the extreme. I therefore requested that I be permitted to adopt the present title.

This is an ambitious enough task, but it allows me latitude in my approach to the subject and some chance to escape pulverization between the millstones of physiology and physical chemistry. My purpose is not to treat the subject in a categorical manner nor to offer solutions for the problems but to approach it philosophically: to present a point of view and some working hypotheses with enough facts to protect them from immediate extinction.

I should like, before approaching the kidney, to mention certain experiments conducted in our laboratory upon the red blood cell, which seem to me to have a bearing upon the whole problem of secretion. This I would define simply as the differentiation of two solutions across a semipermeable membrane. The red blood cells lend themselves to a study of this process because they can be separated from their environment, the plasma, so that the compositions of both media can be determined by direct analysis. All of you are familiar with the striking differences between them. Proteins are more concentrated within the cell; only glucose, urea, and other neutral organic solutes of small molecular size are evenly distributed. Potas-



## CONCENTRATION BY THE KIDNEY

sium is the chief inorganic base in the cells, sodium in the serum; calcium is almost excluded from the cells, which contain more magnesium than the serum. Bicarbonate and chloride are far more concentrated in serum than cells; the cells contain large quantities of organic phosphate esters that are almost entirely lacking in serum. In the test tube, so long as the cells are properly treated, it is impossible to alter their composition by adding the most diverse inorganic solutions. They respond merely by exchanges of water, so regulated that the osmotic pressures on the two sides of the membrane remain always identical. This has been demonstrated by additions to the blood of water and of carbonates, chlorides, sulfates, and phosphates of both sodium and potassium (1). Such experiments have been cited as evidence that the membranes of the red blood cells are impermeable to inorganic bases and to most acids. From indirect evidence of various kinds it has been deduced that most of the cells in the body behave in a similar manner. It is quite obvious, however, that if cells are to function they must have access to the materials essential to their activities, among which belong potassium, magnesium, and phosphate. Their impermeability to these elements must, therefore, be conditioned.

When experiments similar to those just described were attempted *in vivo* by examining the blood of patients before and after the administration of large amounts of hypertonic salt solutions, a peculiar paradox presented itself. Again the volumes of the red blood cells varied directly in proportion to the base of serum. On the other hand, chemical analyses revealed unmistakably that the base in the cells had changed. In the circulating blood in the living subject, then, the impermeability of the cell envelope to base is not absolute. Moreover, base can enter or leave the cell without, apparently, altering osmotic pressure. This suggests that the base may not all be osmotically active. Now there were certain facts that might have led us to anticipate this paradox. In the first place it had been discovered that the correlation between water and base in cells was extremely rough. In the second place the correlation be-

## JOHN P. PETERS

tween the concentrations of base in serum and cells was also poor, the base in the latter being far more variable. In connection with other cells it had been demonstrated by Harrison and Darrow (2) that after adrenalectomy the cells retained or took on potassium without an equivalent amount of water. Nevertheless in this condition the cells responded in the usual manner to changes of the osmotic pressure and base concentrations of serum. It was recognized that potassium, magnesium, inorganic phosphate, and organic phosphate esters played particular individual roles in the metabolic activities of the cells. This suggested that they had certain selective chemical affinities. Magnesium and phosphate esters performed their functions by chemical reactions in which certain proteins participated as coenzymes. It was reasonable to suppose, then, that these inorganic components and proteins formed combinations in the course of the reactions.

The problems which confronted us at this point were: first, to find conditions by which base could be driven across the red blood cell membrane in the test tube; second, to ascertain the state—or, if you will, activity—of the inorganic components of the cells. In neither effort have we been entirely successful, but we have some suggestive results. In describing the *in vitro* experiments above I was careful to state that the red cells, if properly treated, were impervious to bases and to the acids enumerated. Proper treatment consisted of keeping them in the cold for purposes of preservation. Of course this also held metabolic processes in abeyance. With this thought in view Dr. Lena Halpern (3) tested the reactions of the cells at incubator temperature, studying, however, the behavior of inorganic phosphate rather than base. She discovered that although, at refrigerator temperature, no inorganic phosphate could be driven across the membrane in either direction by extreme alterations of its concentration in serum, at incubator temperature the inorganic phosphate derived from hydrolysis of organic esters poured out of the cells even against a concentration gradient produced by the addition of inorganic phosphate to the

## CONCENTRATION BY THE KIDNEY

serum. Obviously the cell membrane was not dead or such concentration gradients would not be maintained. But to make assurance doubly sure it was proved that the current of inorganic phosphate could be reversed by the addition to the system of glucose, which also reversed the direction of the metabolic processes, causing phosphate esters to be reformed at the expense of inorganic phosphate.

These facts are intriguing enough, but what I wish to emphasize is certain of their implications. When this cell was at rest in the cold it acted like a perfectly inert osmometer, impervious to all the substances that give it its specialized properties, responding passively to osmotic changes in its environment by taking on or giving up water. But how could it do otherwise and maintain its integrity without the expenditure of energy? The mere maintenance of differentiation, in the absence of the mysterious restraints that we label impermeability, would be an impossibility without continuous work. When metabolic activity begins it brings with it demands for new materials and changes in the chemical internal environment, but it also provides the energy by which these changes can be effected without dissolution of that environment. Phosphates could move across that red cell membrane, when the cell was active, against the force of osmotic pressure. They were secreted by the cell, presumably by the force engendered by these metabolic activities. However, there is no reason to believe that this energy was provided only to implement the transfer. This latter was, so far as we know, only one feature of the series of linked reactions in the continuing metabolism of the cell.

If, then, we were to investigate this cell, as we have been wont to investigate the kidney, solely in connection with the exchanges of solutes, or more especially a single solute, we need not be surprised to find that these transfers could not be correlated with the energy production of the cells, that they caused no detectable ripples in the smoothly flowing current of continuing oxidations. Is it necessary to believe that, because the kidney cells seem to be more seriously engaged in the business

## JOHN P. PETERS

of transferring water and solutes, they conduct their business in an unorthodox manner? Can we even be sure their transfer business is more active; or does it only seem more active because it is more highly organized? It is time to break down the barrier that has been reared between the physiologists and the physical chemists, to merge the segregated vested interests of the membrane equilibriumites and the metabolites. Proving that cells can lift themselves by their boot straps, that they can perform osmotic feats of incredible magnitude without even a sensible glow or an extra throb of the pulse, is intellectual exercise of the first order; but it seems hardly necessary when there is so much loose energy lying around; it becomes highly unrealistic when it is discovered that these osmotic feats do not occur in the absence of energy.

From a purely mathematical point of view it is impossible to strike an osmolar balance between the components of cells and serum. Apparent success has been achieved only by subconscious or self-conscious neglect of certain cellular constituents. Such incomplete evaluations, stepsons of mechanistic wishes, can only befog the issue. The logical inference must be that certain solutes which are credited with osmotic activity in these calculations are restrained from such activity in the cells. These restraints are dissolved either with the extinction of vital activities or through our clumsy and inept analytical procedures. The formation of glycogen from glucose is a classic example of the osmotic inactivation of a solute, a means by which the cell is enabled to store carbohydrate without imbibing a proportional amount of water. Combinations of calcium with protein in the serum appear to be only scantily dissociated; this calcium, therefore, is osmotically inert. Some of the  $\text{CO}_2$  in red blood cells is now known to exist neither as dissolved  $\text{CO}_2$ , carbonic acid, nor as bicarbonate, but as carbhemoglobin, to the distraction of the Donnan equilibrium into which it had been squeezed with such pains. What is the state of the inorganic solutes in the red blood cells? The literature says that when laked cells are dialyzed against serum or saline the inorganic elements dis-

## CONCENTRATION BY THE KIDNEY

tribute themselves as if they were all free and osmotically active independent agents. But in none of these experiments were any attempts made to preserve the vitality of the blood. In fact, in all of them organic phosphate esters, among other things, had entirely decomposed into their constituent parts.

With the lessons of Dr. Halpern's experiments behind them Miss Hald and Dr. Solomon this year attempted to ultrafilter living laked blood in a Lavietes capsule through a cellophane membrane. To keep the blood alive they laked it by alternate freezing and thawing in liquid air or carbon dioxide and ice

TABLE 1. — AVERAGE RATIOS OF DISTRIBUTION OF SOLUTES BETWEEN THE WATER OF ULTRAFILTRATE AND OF SUBSTRATE

Solute	Treatment of Blood		
	Frozen 7°	Saponin 7°	Frozen 37°
<i>Substrate/Ultrafiltrate Concentration Ratios</i>			
Inorganic phosphate .....	1.3	1.5	1.4
Organic phosphate .....	29.5	1.2	1.0
Sodium .....	1.2	1.0	0.9
Potassium .....	1.7	(2.7)*	1.4

\* In one experiment a ratio of 5.62 was obtained. The other two ratios were 0.98 and 1.51. If the extremely divergent figure is omitted the agreement of averages is far better. Especial difficulty was encountered in the saponin experiments because of the minute quantities of ultrafiltrate that could be secured for analysis.

water respectively, or by adding saponin to chilled blood. It was then transferred to a Lavietes capsule under pressure and stored in the refrigerator. In still another series of experiments ultrafiltration was carried out at incubator temperature. The results of the experiments are summarized in Table 1. In this table are given only the distribution ratios of the various components between the water of ultrafiltrate and of substrate after the hemolyzed blood had stood for eighteen hours at the temperature indicated. If a substance is distributed evenly throughout all the water without restraint its ratio should be 1.0. If it forms dissociable compounds with protein and therefore is controlled by the Gibbs-Donnan equilibrium its ratio should be greater than 1.0 (approximately 1.5) in the case of

## JOHN P. PETERS

the bases Na and K, less than 1.0 (nearer 0.7) in the case of the phosphates.

That the integrity of the blood was not perfectly preserved was evident from the fact that a small fraction of the organic phosphate broke down in the process of hemolysis. The greater portion, however, remained unchanged. In all the experiments inorganic phosphate distributed itself counter to the dictates of the Donnan equilibrium as if a considerable portion was not filterable. In the frozen blood ultrafiltered at 7° the organic acid soluble phosphate remained intact and only a small fraction was filterable. In the saponin blood, also, the organic phosphate was not broken down, but the greater proportion became filterable. In the frozen blood filtered at incubator temperature a large part of the organic phosphate was broken down to form inorganic phosphate, and the remainder became filterable. It appears as if in the saponin and in the incubated frozen bloods phosphate esters distributed themselves evenly between the water of substrate and filtrate. The ratio of sodium, less than 1.0 in the frozen blood at 7°, approached unity under the influence of saponin and at incubator temperature. The parallel behavior of sodium and acid soluble phosphorus suggests that ordinarily both sodium and phosphate esters may be restrained by combination with larger molecules. These combinations seem to be ruptured by saponin and by incubation. In addition the uniform distribution of sodium in the saponin and incubator experiments is quite incompatible with the Donnan equilibrium. The distribution ratio of potassium, on the other hand, was always greater than 1.0, as the Donnan theory demands, and was not consistently and significantly affected by any treatment to which the blood was subjected. The differences between the distributions of potassium and sodium may indicate that hemoglobin combines selectively with potassium, but not with sodium. More important for our purposes than these details, however, is the general principle that the inorganic components of the red blood cell are not freely diffusible even when the restraint imposed by the membrane is removed. This can

## CONCENTRATION BY THE KIDNEY

only mean that they are osmotically inactivated, presumably by incorporation in nondialyzable compounds.

This does not, of course, relieve the cell membrane of responsibility for the segregation of these elements. The effort to reproduce the behavior of living cells by this laked blood and cellophane was not altogether successful. When whole blood is incubated organic phosphate esters are broken down to form inorganic phosphate; the cells swell, indicating that the osmolar concentration within the cells has been increased by the disintegration of chemical compounds; but no organic acid soluble phosphorus, and but little sodium or potassium, crosses the cell membrane. The inactivation of the organic phosphorus serves, therefore, not to hold the phosphate esters within the cell but to prevent them from influencing the osmotic pressure of the cell contents. By this means, presumably, the cells are enabled to vary their load of these useful compounds in conformity with metabolic needs without being forced simultaneously to assume or discharge an equivalent amount of water.

Only now, with much of my time spent, can I turn to the real subject assigned to me, the kidney. And even now I shall hesitate long enough to interpolate one more assumption, or rather suggestion: that the kidney cells may resemble all other cells in the body in maintaining, throughout their diverse activities, an osmotic pressure identical with that of the blood serum and interstitial fluids.

The erroneous concept, initiated by Koranyi, that only concentration involves osmotic work, has unhappily been perpetuated—in the clinic at least. If the work of the kidney per se, in the secretion of urine, as distinguished from the work performed by the circulatory system in forcing blood through the glomerular filter, could be resolved into osmotic components alone, it would not be related to the amount by which the total osmotic pressure of the urine exceeds that of the blood. Instead it would be a function of the volume of urine excreted and the relative concentrations in serum and urine of each individual component, regardless of the direction of the concentration

gradient. It takes just as much work to separate salt from water as it does to separate water from salt. Osmotic work is not measured by the differences of total osmotic pressure between two complex solutions, but by the sum of the differences of the partial osmotic pressures of all the individual osmotically active components of the solutions. This may seem hopelessly elementary to many of you. Nevertheless, its import has been generally neglected in the literature. Such considerations lead to the conclusion that if concentration of the urine is a limiting factor in the formation of urine, the limitation is not imposed by the capacity of the entire kidney to perform work.

Because the loop of Henle is lacking in all animals except birds and mammals, which alone can excrete urine with osmotic pressure greater than that of serum, the power of withdrawing water against osmotic pressure has been attributed to this segment of the renal tubule. If, as has been suggested (5, 6), removal of the posterior lobe of the pituitary, establishing diabetes insipidus, paralyzes this segment, the extremely dilute urine characteristic of this disease must represent the fluid usually presented to the loop of Henle. It is hardly necessary, after what has been already said, to point out that the osmotic work of the kidney does not begin with the loop of Henle. An enormous amount of energy must have been expended on the way to the loop in the absorption of solutes; and the cells of Henle's loop can withdraw their first quota of water with complete conservation of energy. Only after the osmotic pressure of the urine has been raised again to equal that of the serum do their labors begin. How they may perform this work must remain a conjecture until further details of tubular activity are known.

There can be no doubt from the experiments of Gamble (7, 8) and Gilman (9) that osmotic pressure is not the only factor limiting the concentration of urine. The latter found that in dogs it was impossible, by intravenous injections of hypertonic saline, to raise the concentration of sodium chloride in the urine above 300 millimolar, while urea could be concentrated twice as much in osmolar terms. Furthermore, if urea and NaCl



## CONCENTRATION BY THE KIDNEY

were injected simultaneously, neither had any limiting effect on the concentration of the other. The osmotic concentration attainable with the combination was equal to the sum of the concentrations that could be reached by the two solutions given separately. Secretion and reabsorption of the two must be effected by and limited by entirely different forces.

The clearances of urea and certain other organic solutes of relatively small molecular size, such as monosaccharides and disaccharides which cannot be utilized, though always lower than those of purely filtered reference substances, bear a constant relation of the latter under extremely varied conditions. The urea clearance is usually about 60 per cent, the xylose clearance about 90 per cent, of the inulin clearance. The excretion of such a substance would be represented by the equation

$$\frac{F(UV)}{P} = \text{glomerular filtration}$$

in which  $F$  represents a constant characteristic of the particular solute in question. At extremely low urine volumes the urea clearance may fall off somewhat from the inulin clearance. Smith (10) has accepted Rehberg's (11) suggestion that the reabsorption of these substances may be accomplished by "back-diffusion"—that is, they may diffuse back, under no power other than their concentration gradients, as water is withdrawn from the tubular urine. It is implicit in this hypothesis that urea must diffuse more slowly than water. Besides providing for a general correlation between filtration and reabsorption with some distortion of this correlation at extreme rates of urine flow, such a mechanism would permit greater concentration of urea than other solutes by modifying osmotic gradients. If, for example, salts were obstructed by the membrane at the tubular face of the cell, whereas urea could penetrate this membrane at a rate only slightly lower than the rate of passage of water, salt would exert far more osmotic resistance than urea would at this point to reabsorption of water. If this membrane were freely permeable to urea, but not to salt, urea would not compete for water with salt at all. It is conceivable that the

## JOHN P. PETERS

passage of salt is retarded at the renal face, urea at the basement of the cells; or that reabsorption of salt, some withdrawal of water, and back-diffusion of urea occur at different points in the tubule. This does not exhaust the hypotheses that may be evolved to explain the facts. Proper selection of hypotheses can only be made from experiments planned with recognition of these facts and eliminating the preconception that all solutes equally impede the reabsorption of water from the renal tubules.

Glucose represents a class of solutes of particular interest. It has been placed among so-called threshold substances, materials that are filtered through the glomerulus and ordinarily completely reabsorbed by the tubules. As, however, their concentration in the blood or plasma exceeds a certain magnitude, an increasing proportion is supposed to escape reabsorption, appearing in the urine. It has long been recognized, by those who have given the matter most careful consideration, that such a concept is not in keeping with the facts, that reabsorption of glucose cannot be governed only by its concentration in the blood. Govaerts (12) and Shannon (13), having reinvestigated the subject by modern methods, have found that the reabsorption of glucose is not controlled by its concentration in plasma, but by the rate at which it is delivered to the tubular cells. There is an upper limit to the capacity of these cells to reabsorb glucose. When this is exceeded additional glucose is excreted in the urine. In clearance terms the equation for this reaction would be

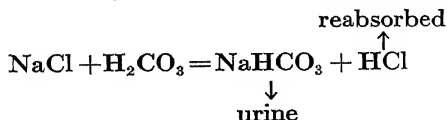
$$\frac{UV + A}{P} = \text{rate of glomerular filtration}$$

where U and P represent concentrations of the substance in urine and plasma respectively, V the volume of urine excreted in a unit of time, and A the rate of reabsorption of the substance under consideration. In the case of glucose, according to Shannon, A=about 200 mg. per min. As I shall point out later this differs strikingly from the equation for the excretion of a true threshold substance. In their behavior toward glucose the tubule cells do not differ greatly from cells in other tissues.

## CONCENTRATION BY THE KIDNEY

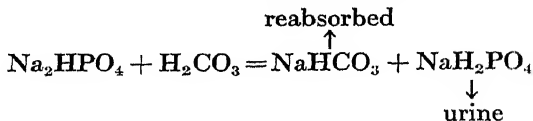
Wierzuchowski (14) has shown that when hypertonic glucose is given to dogs intravenously at an increasing rate, a point is reached beyond which, apparently, the glucose is entirely excluded from the tissue cells. Glucose may enter the tubule cells, as it seems to enter other cells, by simple diffusion, to be disposed of by metabolic processes varying with the speed at which glucose is offered, up to the maximum metabolic capacity of the cell. One need only suppose that the carbohydrate metabolism of those tubule cells which are particularly responsible for glucose absorption is so active that concentrations of glucose within them are always minimal. When the carbohydrate metabolism of muscle is greatly accelerated by insulin, free glucose may disappear from the muscle cell (15) and from the blood flowing through these muscles. After phlorizin glucose is reabsorbed much as other sugars are, as if this drug had poisoned the metabolic processes, leaving nothing but back-diffusion. If the process of reabsorption is, as is generally assumed, conservative, aimed to prevent wastage of a useful nutritive element, the metabolic response cannot involve combustion of all the carbohydrate, but merely temporary storage in an inactive form (possibly as glycogen) and its return to the blood stream as free glucose again. The combustion of a portion of the reabsorbed glucose would provide a source of energy for these reactions. Ralli (16) and Heinemann (17) have found that the excretion of ascorbic acid follows the same principles as the excretion of glucose. Other useful filterable substances may be treated in the same manner.

It has generally been assumed that the excretion of individual ions could be investigated, like that of neutral organic compounds, as if they were independent agents. For instance the process by which the urine is alkalized in the tubules of the kidney has been depicted in the following manner:



# JOHN P. PETERS

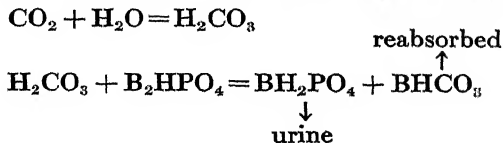
by reactions that could not possibly be reproduced in the test tube. The acidification of urine, which consists of the withdrawal of base from buffer acids, was similarly pictured



How a pH as acid as 5 could be achieved at CO<sub>2</sub>-tensions as low as those of body fluids was, indeed, a little puzzling. Yet measurements by Gamble (15) seemed to have proved that the CO<sub>2</sub>-tensions of blood and urine were identical. The subject took on an entirely different complexion when Mainzer (19) discovered that the CO<sub>2</sub>-tension of the urine might rise as high as 200 mg. Hg, which was substantiated by Van Slyke (20). Apparently the tubule cells are more impervious to CO<sub>2</sub> or carbonic acid than to the bicarbonate ion. Since the reaction of a buffer solution like urine depends on the ratio of carbonic acid to bicarbonate

$$[\text{H}^+] = K \frac{[\text{H}_2\text{CO}_3]}{[\text{NaHCO}_3]}$$

to secure a highly acid urine it is necessary only that bicarbonate reabsorption be accelerated, while reabsorption of CO<sub>2</sub> be retarded. In the defense against alkalosis the ability to retard reabsorption of CO<sub>2</sub> is a distinct advantage because it permits the excretion of larger amounts of bicarbonate with less disturbance of urinary reaction. It may not be inappropriate at this point to speculate about the work involved in the acidification of urine represented in the following reactions:



It should require less osmotic work to concentrate carbonic acid than to concentrate bicarbonate because the former is so

## CONCENTRATION BY THE KIDNEY

slightly ionized. In addition, when carbonic acid reacts with an alkaline salt to form a neutral salt, energy will be produced. The work required for all the reactions would be hard to measure, but would not be proportional merely to the final difference between the osmotic pressures of serum and tubular urine. One is tempted to wonder whether the progressive changes in the proportions of bicarbonate and chloride in the alimentary canal below the pylorus may also depend merely upon variable permeability to  $\text{CO}_2$ .

According to the filtration-reabsorption theory, the volume of glomerular filtrate is probably as much as one hundred times as large as that of the urine actually excreted. There must be in this filtrate such large quantities of all inorganic salts that it is quite unnecessary to postulate incongruous chemical interchanges of ions within the lumen of the urinary tubule. As far as present knowledge goes, except for reactions which can be duplicated in the test tube by changing the reaction of the medium with  $\text{CO}_2$ , inorganic elements can be pictured as either passing through the tubules or being withdrawn from them as inorganic salts, not as independent ions. If this is the case, in the process of reabsorption each ion must be restrained by its mate. It can, therefore, hardly be profitable to study the excretion of individual ions. The erroneous deductions that may be derived from such studies can be illustrated by a few examples.

A patient was given by intravenous injection 250 cc. of hypertonic sodium sulfate solution. The serum was analyzed before and after the injection and the urine excreted during the interval was analyzed for sodium and sulfate. The results of the serum and urine analyses are presented in Table 2. Although equivalent amounts of sodium and sulfate were given, sulfate increased far more than sodium in serum because the hypertonic solution caused the abstraction from the cells of a certain amount of fluid, which diluted the sodium originally in the interstitial fluid and serum. The urine, however, contained almost exactly equivalent amounts of sodium and sulfate. The sulfate clearance estimated by the usual method from the vol-

## JOHN P. PETERS

ume of urine and the concentrations of sulfate in urine and serum,  $\frac{UV}{P}$ , is 4.4. The clearance of sodium by the same method of calculation is only 0.44, because, although the numerator of both equations—that is, the amount excreted—is the same, the denominator, the concentration in the serum, is ten times as large in the case of sodium as in that of sulfate. The agree-

TABLE 2. — INTRAVENOUS INJECTION OF 102 M.EQ. OF  $Na_2SO_4$  AND 250 CC.  $H_2O$ 

	Serum			Urine
	Before	After	Change	In Interval
	<i>m.eq.</i>	<i>m.eq.</i>	<i>m.eq.</i>	<i>m.eq.</i>
$SO_4$ .....	0.5	15.3	14.8	67.4
$Na$ .....	138.8	144.5	5.7	63.0

$$\frac{UV}{P} \text{ (for } SO_4) = \frac{67.4}{15.3} = 4.4$$

$$\frac{UV}{P} \text{ (for } Na) = \frac{63.0}{144.5} = 0.44$$

$$\frac{UV}{P} \text{ (for } Na \text{ increment)} = \frac{63.0}{5.7} = 11.0$$

ment is no better if the increments of the two elements are used. Nevertheless the equivalence of the excretory fractions of sulfate and sodium can best be explained on the ground that the clearance of the moiety of sodium combined with sulfate is identical with the clearance of the sulfate. The compound, sodium sulfate, is being eliminated as such. This particular fraction cannot be distinguished, however, by a study of sodium alone.

Electrolytes have, in general, been placed in the class of threshold substances, although their mode of excretion differs sharply from that of glucose. Figure 1 illustrates results of some studies of the excretion of sulfate after injections of various sulfate salts, carried out by Schwartz, Winkler, and Smith (21). It will be seen that in every instance there appears to be a linear relation between the rate of excretion of sulfate and

## CONCENTRATION BY THE KIDNEY

its concentration in the serum. This is quite independent of the nature of the base with which sulfate is combined, since experiments with magnesium, potassium, and sodium salts are included. In other words this figure can be taken to represent the behavior of the kidney toward exogenous sulfate salts. It should be unnecessary to add, after what has been said earlier, that if the excretion of the associated basic ions were plotted in the same manner, all semblance of uniformity would be lost. It may be noted, furthermore, that if these lines are extrapolated they all seem to intersect the abscissa at about the same point, in the region of 3 m.eq. per liter, which happens to be the concentration of sulfate in normal serum, denoting that excretion of sulfate should cease when its concentration in the serum falls to this point. In terms used for the representation of clearances, the equation for such a reaction should be

$$UV = r(P - a) \text{ or } \frac{UV}{P - a} = r$$

in which  $r$  is a proportionality factor relating excretion rate to serum concentration and  $a$  is a constant representing the threshold—that is, the serum concentration at which sulfate excretion should cease. You will see that this differs radically from the equation describing the excretion of glucose. Sulfate has been placed among the nonthreshold substances, because under no conditions is the urine found free from sulfate. In actual point of fact the thresholds predicted from this figure are only apparent. Endogenous sulfate excretion continues at a slower rate at levels below 3 m.eq. per liter. The lines in Figure 1 describe the excretion of moderate amounts of exogenous sulfate salts.

When larger amounts of sulfate were given and concentrations in the serum were further increased, the direction of the curves, depicted in Figure 2, changed rather strikingly. (Unfortunately only sodium sulfate could be used in these experiments because magnesium and potassium salts are too toxic.) Again the points defining the excretion of sulfate in each of the three experiments, all on the same animal, assume a linear arrange-

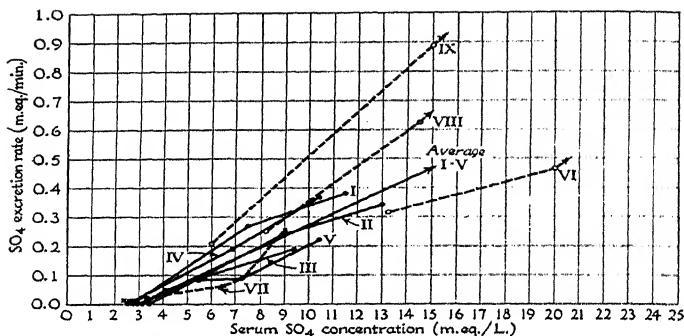


Figure 1. The rate of excretion of sulfate compared with its concentration in serum of dogs after intravenous injections of solutions of the sulfates of magnesium, potassium, and sodium. (Drawn in the Medical Art Shop, University of Minnesota.)

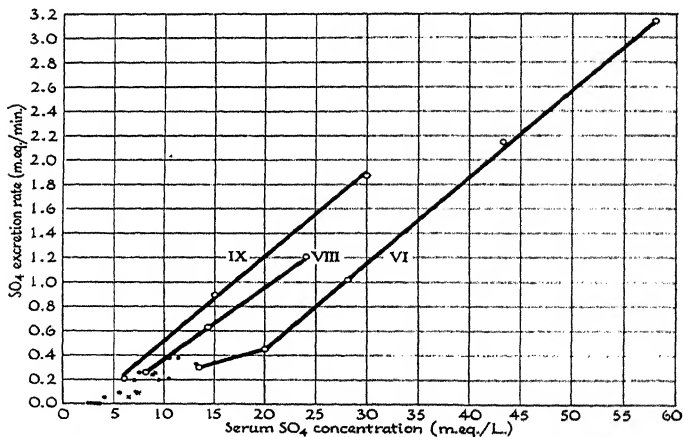


Figure 2. The rate of excretion of sulfate compared with its concentration in serum of dogs after injections of large amounts of sodium sulfate. These curves are included in Figure 1. (Drawn in the Medical Art Shop, University of Minnesota.)



## CONCENTRATION BY THE KIDNEY

ment. But the lines are far steeper, indicating a more rapid rate of excretion in proportion to the serum concentrations, and their apparent intersections with the abscissa indicate far higher thresholds.

These high concentration lines are amazingly straight, considering that during the same experiments creatinine clearances, presumably measures of glomerular filtration in the dog, varied widely. Sulfate clearances were always distinctly below simultaneous creatinine clearances, indicating that above the apparent threshold,  $\alpha$ , the tubules seem always to leave in the tubular urine an amount of sulfate proportional to the concentration of sulfate in the serum, regardless of the quantity delivered to them through the glomerular filtrate. How this can be accomplished is quite incomprehensible and therefore, for the moment, an unprofitable subject for discussion.

The important fact is that as the concentration of exogenous sulfate entering the serum rises, the excess of sulfate is excreted far more rapidly than is the native or endogenous sulfate, and apparently according to a different principle. The rate of excretion of the extra sulfate bears a definite relation to its concentration in the serum. Whether endogenous excesses of sulfate would be treated in the same way is difficult if not impossible to determine, because the breaks in the clearance curves occur at serum concentrations that are never encountered in life unless renal function is profoundly impaired. The temptation is strong to conjecture that the sharp breaks in the direction of the clearance curves are determined not by some critical concentration of sulfate in the serum but by the fact that sulfate in some unusual form has gained access to the tubular urine. The subject takes on more importance in the light of other experiments, which indicate that phosphate, chloride, sodium, potassium, calcium, and magnesium can all be shown to behave in a similar manner if the experimental stage is so set that confusing factors are excluded.

Interactions between inorganic salts of another nature are illustrated by Figure 2. Although the three lines relating excre-

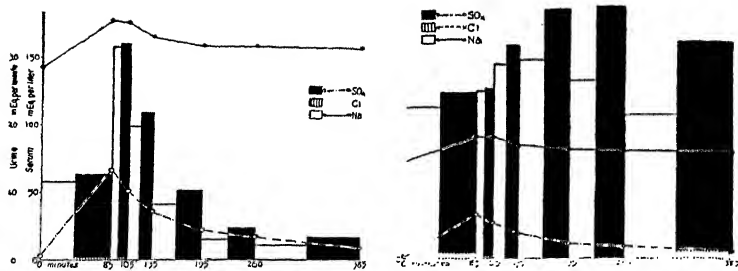


Figure 3 (left). The excretion of sodium, sulfate, and chloride by the dog after intravenous injection of about 1,000 cc. of twice physiologically normal sodium sulfate. The rate of excretion of each substance in every period, represented by columns, is compared with the concentration of that substance in the serum, represented by points and lines, at the beginning of the period. The injection was given during the first period. This experiment depicted by VI, Figure 2.

Figure 4 (right). The concentrations of sodium, chloride, and sulfate in the urine compared with the concentrations of sodium and sulfate in the serum of the dog after intravenous injection of about 1,000 cc. of twice physiologically normal sodium sulfate solution. This is the experiment depicted in Figure 3.

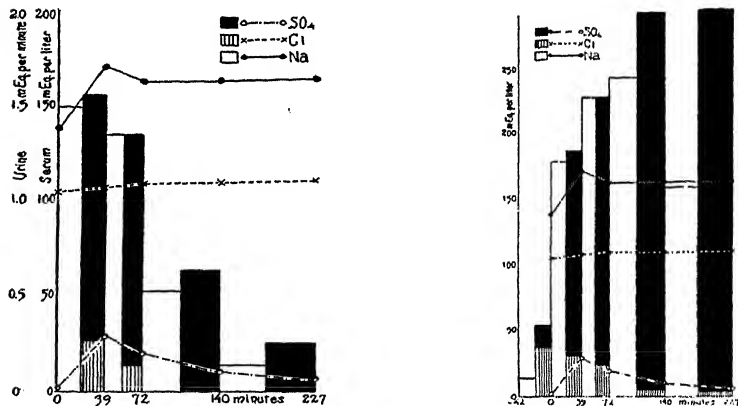


Figure 5 (left). The rate of excretion of sodium, sulfate, and chloride by the dog after intravenous injection of about 1,000 cc. of a solution containing equimolecular concentrations of sodium sulfate and sodium chloride, the combined concentrations yielding a twice physiologically normal solution. The injection was given between 0 and 39 min. This experiment depicted by VIII, Figure 2.

Figure 6 (right). The concentrations of sodium, sulfate, and chloride in the urine, compared with the concentrations of the same substances in the serum of the dog after intravenous injection of about 1,000 cc. of a solution containing equimolecular concentrations of sodium sulfate and sodium chloride, the combined concentrations yielding a twice physiologically normal solution. This is the experiment depicted by Figure 5.

## CONCENTRATION BY THE KIDNEY

tion of sulfate to serum concentration all have the same slopes, they have entirely different positions. There was one recognizable difference in the conduct of the three experiments. The curve to the right illustrates the effect of injecting pure sodium sulfate; the curves to the left illustrate the influence of adding increasing amounts of sodium chloride to the fluid injected. This can only mean that chloride facilitates the excretion of sulfate. In terms of Cushny's theory, it lowers the threshold for sulfate; in more modern terms it diminishes the reabsorption of sulfate. In turn chloride seems to be selectively reabsorbed from the tubules to give preference to sulfate. This inhibition of chloride excretion cannot be attributed merely to dilution of sodium chloride by sulfate. The mutual effects of the ions are far more complicated.

Figure 3 shows the concentrations of sodium and sulfate in the serum and the excretion of sodium, sulfate, and chloride after intravenous injection at 0 minutes of about 1,000 cc. of twice physiologically normal sodium sulfate. The parallelism between the concentration of sulfate in serum and its excretion in the urine are quite patent. It will be seen that after the first two periods chloride practically disappears from the urine. Finally, it may be noted that the amounts of sodium excreted correspond fairly closely to the amounts of sulfate. Figure 4 is similarly constructed, except that concentrations in urine rather than total amounts are given. As the water of the body becomes impoverished by diuresis, the concentration of sulfate in the urine rises even though its concentration in the serum falls. The suppression of chloride is again evident.

Figure 5 shows the effects of the simultaneous injection of a liter of sodium sulfate and sodium chloride of equiosmolar concentration, the combined concentrations yielding a twice physiologically normal solution. During the first three periods the excretion of all three ions—sodium, sulfate, and chloride—diminishes, but that of chloride falls most rapidly. By the end of the second period it has practically disappeared from the urine, although its concentration in the serum is higher

than it was at the beginning of the experiment. Sodium also, after a slight early drop, rises perceptibly in the serum during the last two periods, although its excretion falls. It will be seen from Figure 6, in which the concentrations of ions in the urine are plotted, that the concentrations of sulfate and chloride vary inversely. During the early diuresis both sulfate and chloride are excreted at a rapid rate. As the water available for urine formation diminishes, the concentration in the urine rises. Some preference is given to sulfate from the outset; as the urine concentration increases, chloride is displaced from the kidney entirely. It has already been shown that chloride seems to lower the threshold for sulfate. The disappearance of chloride from the urine when its concentration in the serum is above normal may be termed, if you wish, elevation of the chloride threshold. In the experiment with pure sodium sulfate the excretion of sodium fell somewhat short of the excretion of sulfate. This deficit is even more marked in the present experiment, although there was enough sodium given to cover both chloride and sulfate and although the concentration of sodium remained greatly elevated throughout the experiment. The sodium tended in general to follow the ions with which it was originally united. At the end of the experiment 77 per cent of the sulfate and 75 per cent of the water injected had been recovered, but only 10 per cent of the chloride and 43 per cent of the sodium. The sodium excreted fell about 10 per cent short of the sulfate and chloride combined.

As yet there is no entirely satisfactory explanation for all these conditions that ions seem to put upon one another. The teleological argument that comparative strangers like sulfate are more summarily evicted than familiar natives like chloride has no scientific significance. To speak of chloride as yielding to sulfate in the competition for water may be vivid, but is not accurate description. Chloride did not disappear from the urine at any critical concentration of sulfate or at any critical osmolar concentration of  $\text{Cl} + \text{SO}_4$ . The sum of the concentrations of these two ions in the urine is not constant. In the third ex-

## CONCENTRATION BY THE KIDNEY

periment, which has not been shown in detail, twice as much of both sodium chloride and sodium sulfate was injected in the same volume of water. In this case urinary chloride was not so completely suppressed; but oddly enough, its concentration in the urine fell progressively even after that of sulfate had passed its peak and was also diminishing. At this point, furthermore, serum sulfate had fallen nearly to the normal level, while serum chloride was enormously elevated.

But granted that the concentration of these two competing ions was the limiting factor which forced the kidney to this peculiar selective activity, there is no reason to identify this limiting factor with the capacity of the kidney as a whole to perform osmotic work. In these experiments sulfate was always more concentrated, chloride always more dilute in urine than serum. Work cannot be measured, as I have already stated, by the concentration gradient of sulfate with total neglect of the directionally opposite concentration gradient of chloride.

It may be, of course, that the selective process takes place in the course of the tubules proximal to the loop of Henle and that this specialized loop has no other function than to withdraw the last fraction of water from the mixture, possibly with an admixture of urea and other back-diffusing solutes. In this case concentration above the osmotic pressure of blood might be limited by the capacity of this segment of the tubule to do work. The reaction of the kidney in diabetes insipidus, especially the relatively negligible effect which this disease has upon the excretion of solutes, argues strongly for such an allocation of functions. If this view should prevail, the behavior of the more proximal tubules in reabsorbing chloride to the neglect of sulfate would be quite comparable to the behavior of the gut, which has been so admirably elucidated by Ingraham and Vischer (22). As yet there have been few rigorous attempts to test this theory; it has been too much a conventional assumption. Gamble's (7, 8) and Gilman's (9) experiments suggest that the same limits are not imposed on concentration by urea as by salt. Experiments must be extended into more dimensions and

## JOHN P. PETERS

must include more variables. It may prove that the work of the loop of Henle is measured and limited not by the total osmotic pressure of the urine but by the extent to which the concentration of solutes that do not diffuse back, or even only inorganic solutes in the urine, exceeds that of the serum.

In any case the over-all work of the kidney, which is all that we can hope to measure, will be determined by far more numerous and complex factors, the nature of some of which I have attempted to analyze. There should be no reason for surprise, then, if the oxygen consumption of the kidney is not accelerated by every excretory increment; nor controversy if two should disagree over the effects on caloric expenditure of any particular diuretic measure: perhaps their respective experimental stages were differently set. Recently the excretion of sulfate was described by Goudsmit, Power, and Bollman (23) in terms other than those you have heard today. The differences in fact are actually slight, but these observers maintained diuresis throughout their experiments by injections of hypertonic sodium chloride, thus distorting the picture.

My aim has been to point out the interdependence of ions, their mutual obligations to one another and to water, and the necessity of approaching the problem of their excretion with a consciousness of the restraints these associations place upon them. In addition I have tried to point out that the kidney must be looked upon not as a mere drain but as an aggregation of cells, which possess, besides their specialized properties, the general attributes of other tissues, including continuing metabolic processes. Finally, I should end on the note on which I began. Studies confined to resting or equilibrium states or studies of inert systems cannot be expected ultimately to solve the problems posed by systems charged with energy. This is accepted without question when the problems involve such matters as muscular work, because our habits of mind have accustomed us to connect muscular work with heat production. But it must be equally true of transfers of solutes across membranes when these involve the production of concentration gradients.

## CONCENTRATION BY THE KIDNEY

Can it not be hoped that some light may be thrown on the mechanism by which these gradients are produced by a study of the chemical and metabolic reactions with which they are linked?

### *References*

1. Peters, J. P. 1938. Transfers of water and solutes in the body. *Bull. N. Y. Acad. Med.* 14:299.
2. Harrison, H. E., and Darrow, D. C. 1938. The distribution of body water and electrolyte in adrenal insufficiency. *Jour. Clin. Invest.* 17:77.
3. Halpern, L. 1936. The transfer of inorganic phosphorus across the red blood cell membrane. *Jour. Biol. Chem.* 114:747.
4. Solomon, R. Z., Hald, P. M., and Peters, J. P. 1940. The state of the inorganic components of human red blood cells. *Jour. Biol. Chem.* 132:723.
5. Burgess, W. W., Harvey, A. M., and Marshall, E. K., Jr. 1933. The site of the antidiuretic action of pituitary extract. *Jour. Pharmacol. Exper. Therap.* 49:237.
6. Robert, F. 1932. Ueber die Einwirkung von Hypophysin und seinen Fraktionen auf den Wasser-Salzstoffwechsel. *Arch. f. exper. Path. u. Pharmacol.* 164:367.
7. Gamble, J. L., McKhann, C. F., Butler, A. M., and Tuthill, E. 1934. An economy of water in renal function referable to urea. *Am. Jour. Physiol.* 109:139.
8. Gamble, J. L., Putnam, M. C., and McKhann, C. F. 1929. The optimal water requirement in renal function. I. Measurements of water drinking by rats according to increments of urea and of several salts in the food. *Am. Jour. Physiol.* 88:571.
9. Gilman, A., and Kidd, N. E. 1938. The osmotic work of the kidney following the injection of hypertonic NaCl, urea and their combination. *Proc. Am. Phys. Soc.* p. 77.
10. Smith, H. W. 1937. *The Physiology of the Kidney*. Oxford Univ. Press, New York.
11. Holten, C., and Rehberg, P. B. 1931. Studies on the pathological function of the kidneys in renal diseases, especially in Bright's disease. I. *Acta med. Scand.* 74:479.
12. Govaerts, P., and Cambier, P. 1934. Elimination comparée de la créatine et du glucose chez les diabétiques. *Acta med. Scand.* 83:317.
13. Shannon, J. A., and Fisher, S. 1938. The renal tubular reabsorption of glucose in the normal dog. *Am. Jour. Physiol.* 122:765.
14. Wierzechowski, M. 1936. The limiting rate of assimilation of glucose introduced intravenously at constant speed in the resting dog. *Jour. Physiol.* 87:311.
15. Cori, G. T., Closs, J. O., and Cori, C. F. 1933. Fermentable sugar in heart and skeletal muscle. *Jour. Biol. Chem.* 103:13.

## JOHN P. PETERS

16. Ralli, E. P., Friedman, G. J., and Rubin, S. H. 1938. The mechanism of the excretion of vitamin C by the human kidney. *Jour. Clin. Invest.* 17:765.
17. Heinemann, M. 1938. The distribution of ascorbic acid between cells and serum in relation to its urinary excretion. *Jour. Clin. Invest.* 17:751.
18. Gamble, J. L. 1922. Carbonic acid and bicarbonate in urine. *Jour. Biol. Chem.* 51:295.
19. Mainzer, F., and Shen, C. T. 1929. Die Bestimmung der Gasspannungen im Harn und anderen carbonathaltigen Flüssigkeiten (zugleich ein Beitrag zur Theorie der  $\text{CO}_2$ -Bindungskurve). *Arch. f. d. ges. Physiol.* 222:1.
20. Sendroy, J., Jr., Seelig, S., and Van Slyke, D. D. 1934. Studies of acidosis. XXIII. The carbon dioxide tension and acid-base balance of human urine. *Jour. Biol. Chem.* 106:479.
21. Schwartz, B. M., Winkler, A. W., and Smith, P. K. Unpublished studies.
22. Ingraham, R. C., and Visscher, M. B. 1936. The production of chloride-free solutions by the action of the intestinal epithelium. *Am. Jour. Physiol.* 114:676.
23. Goudsmit, A., Jr., Power, M. H., and Bollman, J. L. 1939. The excretion of sulfates by the dog. *Am. Jour. Physiol.* 125:506.



## **Part II. Some Recent Investigations in Metabolism**



# ORGANIC CHEMISTRY IN THE PURSUIT OF VITAMIN RESEARCH

BY

Lee Irvin Smith

*Professor and Chief, Division of Organic  
Chemistry, University of Minnesota*

IN THE field of vitamins and hormones we deal with substances which exert profound biological effects by their presence in very minute amounts. It may therefore be said with some reason that in this field we approach rather closely the idea, advanced by Berzelius many years ago, that life phenomena are essentially catalytic. Be that as it may, it is true that there exists a large number of substances—called variously vitamins, hormones, and enzymes—whose action is very much akin to the action of catalysts; and these terms—vitamins, hormones, and enzymes—while fairly definite in themselves, might logically be regarded as special classes of a very large group of substances which have as yet no widely accepted name but which might be called biological catalysts. As yet all of the substances clearly belonging in these groups are organic in nature, but it is quite within the bounds of possibility that there exist inorganic substances which exert profound effects in such small amounts that they might almost be called *inorganic vitamins*. Indeed substances of this nature are already known in connection with both plant and animal metabolism.

The problem, then, involved in vitamin research is essentially to discover, by means of carefully controlled experiments, the nature of a substance which is causing a given effect by its presence in extremely small amounts. And since the active substance is organic in nature and its effects are biological, it requires the combined efforts of biologists and chemists to solve the problem. In no other field, perhaps, has there been such an absolute necessity for cooperative efforts among different

## LEE IRVIN SMITH

branches of scientific research as in this field of vitamins and hormones. Since vitamins are obtained largely from foods, the discovery of the existence of a vitamin has usually come about as the result of carefully controlled feeding experiments; indeed our knowledge of vitamins today may be counted largely as one of the triumphs of the modern techniques in the field of nutrition.

The study and elucidation of the nature of a vitamin and of its biological action may be divided roughly into five stages:

1. Recognition of a biological effect not caused by any functional abnormality, which is traced to a lack of something in the "highly purified" diet used, because the biological effect does not result when a well-balanced diet made up of naturally occurring, unprocessed foodstuffs is used.

2. By a process of selective elimination the substance responsible for the biological effect is located in one or more of the dietary components, and a foodstuff is found which is shown to be particularly rich in the unknown substance. Here we have experimental production of the disease state, which can be cured by administration of the active dietary factor. The material found to be rich in the unknown substance is processed in various ways, by extraction, partition, and the like, each of the components—proteins, fats, nonsaponifiable part, etc.—being tested experimentally for its curative action in the disease state. In this way it is discovered that the active factor is concentrated in one or two of the extracts from the foodstuff. And at this point comes the first clue, usually by elimination, as to what sort of a chemical substance the active principle may be. For instance, if it is found that the active principle becomes concentrated in the nonsaponifiable, non-nitrogenous part of the extracts, it follows that the active principle cannot be a fat (ester) nor a protein. Along with this second stage in the study of the vitamin come improvements in the methods of bio-assay; these usually lead, at about this stage, to well-developed, reliable methods which are as simple as is possible under the circumstances.

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

3. At the third stage the entrance of physical and chemical methods begins, and these methods play an ever increasing role as the study develops. Some physical criterion is sought by means of which the presence of the active substance may be determined quickly, without waiting for the results of the time-consuming bio-assays. A very common criterion is the absorption spectrum; often this will show a maximum or minimum at a given point, which correlates with biological activity and so is characteristic of the active substance. The concentrates are worked over and further separated into fractions, one or two of which will show greatly increased biological activity, and the others less than the original concentrates. Soon very active concentrates are obtained, and with these, *chemical* tests are made to determine any active groups that may be present—ketone, aldehyde, hydroxyl, and the like—and to find if possible a reagent which will precipitate the active substance. In any event the active principle, which usually long before this stage has been recognized as a vitamin and given a letter in the vitamin alphabet, is finally isolated in pure form and its composition established by analysis. By the time this stage is reached thousands of bio-assays will have been made, and often much may be known, in a general way, about the substance chemically and biologically.

4. In the fourth stage the methods of organic chemistry are applied to the substance to determine the structure and then to synthesize it in the laboratory. This is usually difficult, because for the most part the vitamins are very complex organic molecules and available only in minute amounts. The chemical degradation of such very small amounts of pure vitamins as are usually available initially in researches of this type is a truly formidable task, and were it not for the modern techniques enabling the chemist to work efficiently with milligrams of material rather than with grams, the task would be quite hopeless. But in this fourth stage the structure is determined and the compound is synthesized. The synthesis always makes available relatively large amounts of the pure vitamin, hence

## LEE IRVIN SMITH

experiments may now be made with potent, standardized, and uniform material at a fraction of the previous cost.

5. The fifth stage, the final one, may be said to be practically endless. The researches at this stage may be subdivided into a) the chemical and b) the clinical.

a) The chemical part of the fifth stage deals with the preparation of analogs of the vitamin in the hope of finding simpler, more readily accessible compounds with high activity. Thousands of new compounds, related structurally to the vitamin, are made; all are tested biologically and any activity is correlated, if possible, with structure. In this way it may be possible to locate the part of the molecule largely responsible for the activity, and having located it, to modify and improve it. If this should be successful a "synthetic vitamin," in the strictest sense of the words, would result. Along with the search for new and simpler compounds of high activity goes, in collaboration with the physiologist, a search for the chemical mechanism whereby the vitamin operates in the body; as examples of active compounds accumulate, certain structural features common to them may enable some possible mechanisms of action—such as oxidation-reduction, etc.—to be excluded while other mechanisms remain still possible.

b) The clinical part of the fifth stage has to do with the use of the vitamin in the treatment of human disease. Nor must one at this stage neglect the veterinary field, for frequently vitamins have found great use in the treatment of domestic animals: this promises to be one of the very important fields of application for vitamin E. From this it is easy to see that vitamins may well be of tremendous importance in the economy of a country and in public health.

In these stages of the study of a vitamin, the first three are the most time consuming. In the beginning of the study of the vitamins a long time elapsed between establishing the existence of a vitamin and isolating it in pure form from natural sources; once the isolation was achieved, the synthesis usually followed fairly quickly. The great amount of time used for the

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

early work was due to a lack of knowledge of procedures and techniques. These had to be developed as the study progressed, and guiding principles had to be established; the methods of nutrition research—controlled diets and the like—also had to be perfected. But with each succeeding study of a vitamin the time intervals between stages have decreased, so that in the most recent vitamin study, that of K, the first four stages were completed in three years. Table 1 shows this trend (1).

TABLE 1. — DISCOVERY OF THE VITAMINS

Vitamin	Existence Established	Isolation	Synthesis	Designation
A .....	1912-15	1928	1937	Vitamin A
B .....	1912-15			
B <sub>1</sub> .....	1912-15	1926	1936	Thiamin
B <sub>2</sub> .....	1933-	1933	1935	Riboflavin
P-P .....	1925-26	1937	1867	Nicotinic acid amide
B <sub>6</sub> .....	1936	1938	1939	Vitamin B <sub>6</sub>
C .....	1919	1932	1933	Ascorbic acid; cevitamic acid
D .....	1922	1927	1927*	Calciferol
E .....	1922	1936	1938	Tocopherol ( $\alpha$ , $\beta$ , $\gamma$ )
K <sub>1</sub> .....	1936	1939	1939	Vitamin K <sub>1</sub>
K <sub>2</sub> .....	1939	1939		Vitamin K <sub>2</sub>

\* Not a true synthesis from simpler substances. As the organic chemist understands the term, vitamin D has not yet been synthesized, although its structure is known with a great deal of certainty.

These stages in the investigation of the nature of a vitamin will now be outlined, using vitamin E (tocopherol) as the example.

*First stage: recognition of the biological effect.*—In 1922 Evans and his collaborators at California (2) described the results of a long series of experiments, which indicated that there was required in animal nutrition a dietary constituent necessary for normal reproduction. Young rats, fed a sufficiently long time on a diet of purified foods with addition of the necessary salts and all of the known vitamins, lost the ability to reproduce. By adding certain vegetable products to the diet the reproductive ability was regained. It followed that there existed, in the added vegetable products, an unknown factor necessary

## LEE IRVIN SMITH

for the normal reproductive ability of rats. Evans designated this as factor-X; later he called it vitamin E. It has also been called the antisterility factor or the reproductive vitamin. Although other vitamins, especially vitamin A, appear to exert an influence on the reproductive ability, this loss is most characteristic of a lack of vitamin E. The existence of vitamin E was at first disputed by several workers, but as the studies progressed it was shown that these workers had used diets not quite free from vitamin E and soon there was general agreement that the factor actually existed. Almost simultaneously with Evans' publications Sure (3) and Mattill (4) published the results of their experiments, which also indicated the existence of the antisterility factor. These results have since been duplicated in many other laboratories.

*Second stage: location of foodstuffs rich in vitamin E, and experimental production of the disease state.*—Further investigation of vitamin E was intensively undertaken by Evans and his associates. Extended series of experiments, involving many thousands of experimental animals, were carried out (5). The results showed that wheat-germ oil was the richest source of vitamin E, but that considerable amounts of the vitamin were also found in cottonseed oil, lettuce oil, rice-germ oil, and other seed-germ oils. The vitamin remains in the unsaponifiable part of the lipid fraction. By processes of partition between different solvents, a sterol-free concentrate was obtained, which was active in single doses of 20 mg. (5).

The characteristic symptoms of lack of vitamin E differ in the sexes. In the female rat (5) normal conception occurs, but this is followed by "resorption sterility." There is the usual pregnancy increase in weight for about 10 days; then the weight decreases, becoming normal at about the twentieth day. No litter is cast. The litter has been resorbed, but the resorption has no effect upon the next oestrus cycle. If a female known to be in this state of resorption sterility is again mated, conception occurs as before. A day or so later the animal is given in the food the substance to be tested. If this is active, the



## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

pregnancy will be terminated by the birth of a litter of living young. The vitamin E activity is usually expressed as milligrams of the substance, fed in a single dose, necessary to cure the sterility and to produce litters in 50 per cent of the animals used (5, 6, 7).

In male animals the characteristic symptoms of lack of the vitamin are associated with the germinal epithelia and the spermatozoa. These degenerate until all sexual power is lost. These changes can be arrested by vitamin E only in the early stages; once the degeneration in the male animal progresses very far, administration of the vitamin is of no use.

Along with these changes in the reproductive organs go other, more obscure degenerative changes elsewhere. Recently Shimotori, Emerson, and Evans (8) have reported on cases of muscular dystrophy caused by lack of vitamin E, and there are growth effects (9) clearly discernible, as well as a characteristic paralysis of the hind quarters (10). Other effects, especially connected with the hypophysis and with the occurrence and growth of tumors, have been reported; but there is not complete agreement, as yet, about the connection between these effects and vitamin E.

*Third stage: isolation.*—At the close of the second stage in the study of vitamin E it was possible to obtain a concentrate from wheat-germ oil which showed activity in doses of 20 mg. These were yellow to red oils which were extremely difficult to concentrate further. By high vacuum distillation Olcott and Mattill (11) were able to obtain a fraction boiling at 200°–250° under 0.05–0.1 mm. pressure, which was active in doses of 5 mg., but the vitamin was damaged in this process by the high temperature necessary for the distillation. Evans subjected the concentrates to partition between petroleum ether and methanol and obtained highly active preparations; Drummond used chromatographic adsorption to achieve the same end. But none of these procedures yielded the vitamin in crystalline form. At each stage in these separations the various fractions were assayed biologically, and also at this time measurement

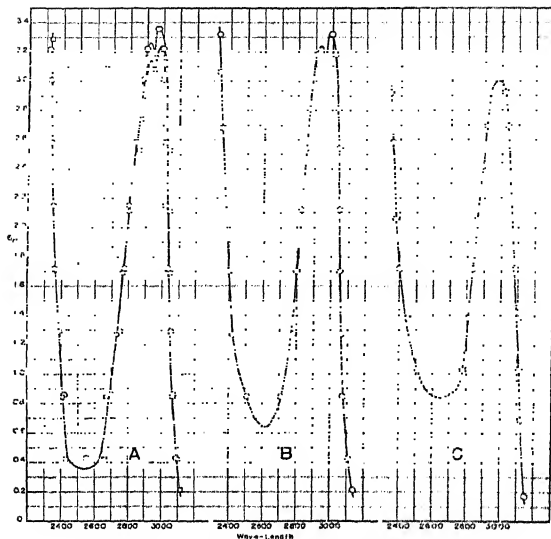
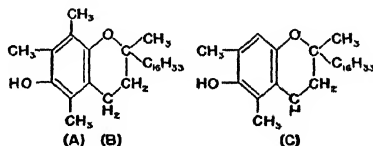


Figure 1. Absorption spectra of tocopherols.



of the ultraviolet absorption spectra of these concentrates was begun. It was found that a parallel existed between the activity and the height of an absorption band at  $2940\text{\AA}$  (12, 13, 54), and this proved to be a reliable guide in following the process of concentration. The curves are given in Figure 1 (54). A is the curve for natural  $\alpha$ -tocopherol, circles and squares representing two different preparations. B is the curve for synthetic *dl*- $\alpha$ -tocopherol, and C is the curve for *m*-xylotocopherol. In Figure 2 are given, for comparison, the curves of three model substances related in structure to the tocopherols; the similar-

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

ties as well as the differences of the chroman and coumaran types are apparent from these curves.

These vitamin E concentrates are readily soluble in all lipid solvents, and only slightly soluble in water. They withstand a temperature of about  $200^{\circ}$  and are fairly stable in the air when in mass, although when finely divided they are attacked by air and lose their activity. Ultraviolet light quickly destroys all of the activity. The concentrates are fairly stable toward acids, much less so toward alkalis. They are resistant to reduc-

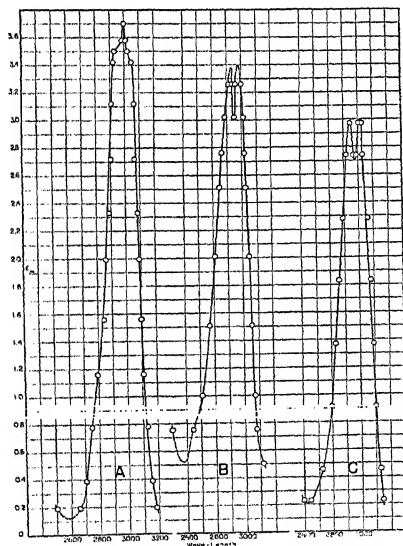
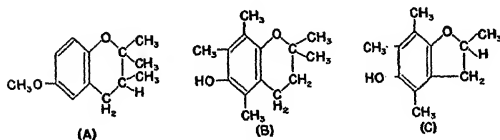


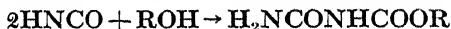
Figure 2. Absorption spectra of simple chromans and coumarans.



## LEE IRVIN SMITH

tion but are quickly attacked by oxidizing agents, even by such mild oxidizing agents as ferric chloride. Acetylchloride and benzoyl chloride react to produce esters, and these esters have practically the same activity as the original material. By comparing the shift in the maxima of the absorption spectra that takes place when phenol is acetylated with that occurring when vitamin E concentrates are acetylated, John (14) was able to deduce that the hydroxyl group in vitamin E was phenolic in nature.

However, esterification of these concentrates by various acids failed to produce solid esters, and it was not until Emerson (15) treated the concentrates with cyanic acid that a solid derivative of the vitamin was obtained. This reaction, characteristic of the hydroxyl group, leads to esters known as allophanates.



By careful purification of the solid obtained in this way from wheat-germ oil concentrates there was obtained first an allophanate melting at  $159^{\circ}$ – $160^{\circ}$  and then a second allophanate melting at  $138^{\circ}$ . These allophanates were hydrolyzed, and each yielded a pale yellow oil. These oils were both highly active, the first in 3 mg. doses, the second in 8 mg. doses. For these individual vitamin E factors Evans coined the name tocopherol; the tocopherols were then designated as  $\alpha$ - and  $\beta$ -tocopherols. From one kilogram of wheat-germ oil, about one gram of  $\alpha$ -tocopherol allophanate may be obtained, although the yield is often much less than this.

$\alpha$ -Tocopherol possesses all of the properties of the highly active concentrates from wheat-germ oil. It shows the same solubility behavior, and the absorption band at  $2940\text{\AA}$  is the same. Analysis shows the composition to be  $\text{C}_{20}\text{H}_{50}\text{O}_2$ . The homogeneity of the preparation was shown by converting it into a solid p-nitrophenyl urethane and a solid p-nitrobenzoate and transforming these into allophanates with the same melting point as that possessed by the original allophanate from the concentrates.

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

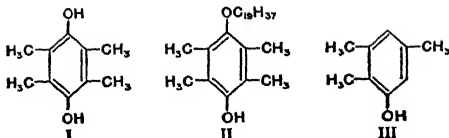
$\beta$ -Tocopherol, obtained in the same way from its allophanate, is likewise an oil. Its properties are almost identical with those of  $\alpha$ -tocopherol, but its composition is  $C_{28}H_{48}O_2$  and so it is a lower homolog of  $\alpha$ -tocopherol. The yield of  $\beta$ -tocopherol from wheat-germ oil is usually much smaller than the yield of  $\alpha$ -tocopherol, but often, from oils of different sources, normal amounts of  $\beta$ -tocopherol can be isolated while almost no  $\alpha$ -tocopherol can be found.

A third allophanate, melting at  $138^\circ$ , has been isolated from cottonseed oil by Emerson, Evans, Olcott, and their associates. This has been named  $\gamma$ -tocopherol allophanate.  $\gamma$ -Tocopherol is likewise an oil, active in 8 mg. doses, and is an isomer of  $\beta$ -tocopherol, having the composition  $C_{28}H_{48}O_2$ .

We have, then, three antisterility factors which are responsible for vitamin E activity. These three tocopherols appear to be the only substances isolated from natural material which certainly possess vitamin E activity, for reports of still other active principles have not been substantiated (90, 20b).

*Fourth stage: determination of the structure, and synthesis.*  
 $\alpha$ -Tocopherol.—As mentioned above,  $\alpha$ -tocopherol possesses the composition  $C_{29}H_{50}O_2$ , which is very close to that of some of the sterols, sitosterol for instance, having the composition  $C_{29}H_{50}O$ . As dehydrogenation with selenium had been of such great value in connection with structure studies in the field of the sterols, it was natural that this method should be applied to  $\alpha$ -tocopherol. McArthur and Watson (16) heated  $\alpha$ -tocopherol with selenium; the result was a yellow sublimate, duroquinone, and a red oil. Somewhat later Fernholz (17) pyrolyzed  $\alpha$ -tocopherol at  $350^\circ$  in the absence of any dehydrogenating agent. There was obtained a good yield of a white crystalline sublimate, identified as durohydroquinone (I), together with a red oil. The simplest assumption which would account for these decomposition products was that  $\alpha$ -tocopherol was a monoether of hydroduroquinone, such as II (in which the group  $C_{19}H_{37}$  contained one saturated ring), for it was known that many alkyl ethers of phenols were cleaved by pyrolysis into

# LEE IRVIN SMITH

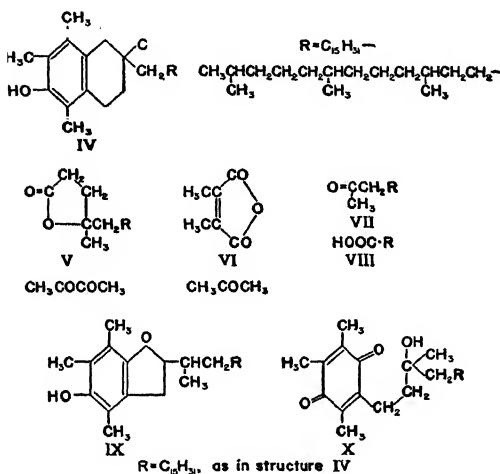


the phenol and an unsaturated hydrocarbon. Accordingly in several laboratories mono-ethers of hydrodurequinone and of other hydroquinones were synthesized. Some of these showed activity when assayed biologically, but these ethers differed markedly from  $\alpha$ -tocopherol in chemical properties, and their ultraviolet absorption spectra were also quite different from that of the vitamin. As a result of these studies it quickly became apparent that  $\alpha$ -tocopherol could not be a simple mono-ether of hydrodurequinone. John, Dietzel, and Günther (14) had also obtained pseudocumenol-6 (iso pseudocumenol) III, by heating  $\alpha$ -tocopherol with hydriodic acid; this result was also difficult to reconcile with the assumption that  $\alpha$ -tocopherol was a simple mono-ether of hydrodurequinone, but it could be reconciled with the assumption that a second ring was condensed with the aromatic nucleus, probably involving an oxygen atom. Bergel, Todd, and Work (18) found that  $\alpha$ -tocopherol, when energetically hydrogenated, absorbed four moles of hydrogen; and they too supposed that an oxide ring was a part of the structure of the vitamin.

The correct structure for  $\alpha$ -tocopherol (IV) was proposed by Fernholz (19) as a result of oxidative degradation, using chromic acid as the oxidizing agent. The products were a  $C_{21}$  lactone (V), dimethylmaleic anhydride (VI), a  $C_{18}$  ketone (VII), a  $C_{16}$  acid (VIII), together with diacetyl and acetone. The hydroxy acid corresponding to the lactone V was transformed into the lactone with extreme ease, indicating that it was a  $\gamma$ -hydroxy acid; moreover, the hydroxyl group of the acid could not be oxidized to a carbonyl group, and was esterified only with difficulty. These facts indicated that the hydroxyl group was tertiary. The  $C_{16}$  acid VIII, when analyzed for  $C-CH_3$  groups, showed three such groups. The structure for

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

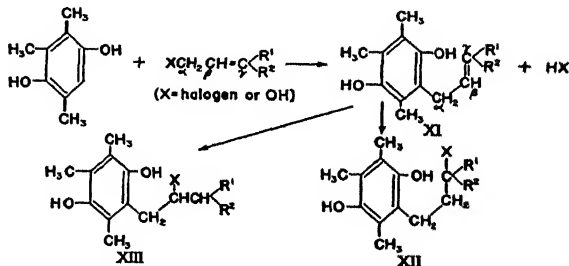
the lactone V can only be written as shown in order to explain the formation from it of a C<sub>18</sub> ketone and a C<sub>16</sub> acid, and when these degradation products are assembled they lead unequivocally to the structure IV, a chroman, for  $\alpha$ -tocopherol. These results do not, of course, lead to the structure shown for the group R, C<sub>15</sub>H<sub>31</sub>. The structure for this group was written on the basis of the C-methyl determination and the experiences gained in other fields of natural products, which frequently contain chains of "isoprene" units joined head to tail.



Karrer (20), although considering both the chroman (IV) and the coumaran (IX) structures for  $\alpha$ -tocopherol, at first preferred the latter. However, John and his associates (21) showed that  $\alpha$ -tocopherol, when oxidized carefully with silver nitrate or ferric chloride, gave a yellow quinone X. This quinone could be reduced to a hydroquinone, the di-p-bromobenzoate of which was quite stable toward chromic oxide, a fact which indicated that the hydroxyl group in X was tertiary. This could only be true if the oxygen ring in  $\alpha$ -tocopherol were a chroman, for the coumaran IX would on oxidation give a hydroxy qui-

none whose hydroxyl group would be secondary and thus susceptible to ready oxidation by chromic oxide. Karrer based his selection of the coumaran formula IX upon the fact that allyl bromide, when condensed with trimethylhydroquinone, does give a coumaran, and when he synthesized  $\alpha$ -tocopherol from phytol bromide, trimethylhydroquinone, and zinc chloride (22) he stated that structure IX was "highly probable."

The synthesis of chromans such as IV is rather simple and easy. The starting materials are hydroquinones (or phenols) having vacant one position in the ring ortho to the hydroxy group (67, 96, 99). These are condensed with allylic halides or alcohols, or with conjugated dienes. Frequently the reaction proceeds so smoothly that neither solvent nor catalyst is required, especially when allylic bromides or chlorides are used. When the alcohols or the dienes are used, it is customary to employ both a catalyst and a solvent. But in any event, because of the great reactivity of the allylic compounds, coupled with the enhanced activity of the aromatic nucleus in polyalkyl benzene derivatives, reactions between the two classes of compounds take place readily, and the products are produced in good yields.

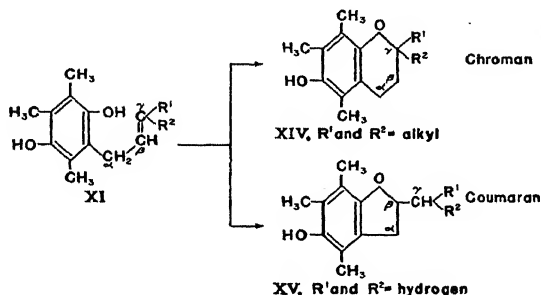


Using the halides or the alcohols, the first step in the reaction appears to be a direct introduction of the allyl group (23, 24) without rearrangement, to give XI. Frequently, when X is halogen, the HX addition products XII or XIII of the allylic compounds XI can be isolated. The second step in the reaction,



## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

the ring closure, involves the addition of the hydroxyl group to the double bond in the side chain of XI, in accordance with Markownikoff's rule. Hence, whether a chroman or a coumaran will be formed in this reaction will depend upon the nature of the groups or atoms attached to the  $\gamma$ -carbon atom in the allylic compound. If these groups are both alkyl, the oxygen of

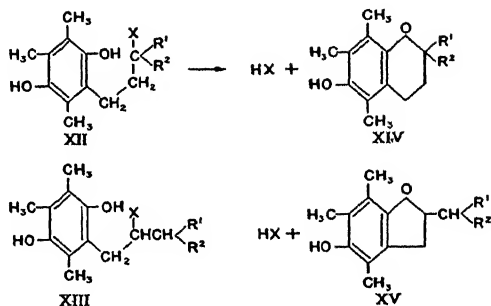


the hydroxyl group will add to the  $\gamma$ -carbon atom, and the product will be a chroman (XIV); while if these groups are both hydrogen, addition will occur in the reverse manner and the product will be a coumaran (XV). When one of the groups is alkyl and the other is hydrogen, the product might be either the chroman or the coumaran or a mixture of the two, although in most of these cases which have been studied so far it is largely the chroman. In a recent paper Karrer, Escher, and Rentschler (102) have made similar generalizations about these ring closures; they have isolated, as condensation products of trimethylhydroquinone and crotyl bromide, *both* the chroman XXXIII and the coumaran XXXIIIa. The structure of the latter was proved by an independent synthesis, using the sequence of reactions shown for the synthesis of XVI, substituting propionyl acetic ester for acetoacetic ester.

The halogen-containing products, XII and XIII, follow the same general rules. These can be readily cyclized to ring compounds, HX being eliminated between the halogen atom and the hydrogen atom of the hydroxyl group. It is to be noted that

# LEE IRVIN SMITH

Markownikoff's rule also plays a part in these reactions, for although XII cyclizes to XIV ( $R^1$  and  $R^2$  are alkyl), the addition of HX to the double bond in XI could occur in two ways, and the mode of addition will be governed by the rule. Thus when  $R^1$  and  $R^2$  in XI are alkyl groups, HX will add so as to produce XII; but when  $R^1$  and  $R^2$  are hydrogen atoms, HX will add so as to produce XIII. The ensuing ring closure by elimination of HX would then give the chroman XIV from XII and the coumaran XV from XIII.



The generalities stated above regarding these reactions were carefully checked by means of model experiments upon simple compounds, the structures of which could be proved by independent syntheses. Thus, when allyl bromide or chloride is condensed with trimethylhydroquinone, the product is the coumaran XVI (24, 26), which is also produced by reduction of the coumaron XVII, whose structure had previously been proved (27). When  $\gamma, \gamma$ -dimethyl allyl bromide is used, the ring closure occurs in the reverse direction and a chroman XVIII is produced (24, 26, 28). The structure of this chroman also was proved by an independent synthesis from coumarin derivatives (XIX, XX, and XXI) of known structure (29, 30). This same chroman XVIII has also been synthesized in other ways (31, 32, 33) so there can be no doubt as to its structure; it is also the product of the reaction between trimethylhydro-

# ORGANIC CHEMISTRY IN THE PURSUIT OF VITAMIN RESEARCH

BY

Lee Irvin Smith

*Professor and Chief, Division of Organic  
Chemistry, University of Minnesota*

IN THE field of vitamins and hormones we deal with substances which exert profound biological effects by their presence in very minute amounts. It may therefore be said with some reason that in this field we approach rather closely the idea, advanced by Berzelius many years ago, that life phenomena are essentially catalytic. Be that as it may, it is true that there exists a large number of substances—called variously vitamins, hormones, and enzymes—whose action is very much akin to the action of catalysts; and these terms—vitamins, hormones, and enzymes—while fairly definite in themselves, might logically be regarded as special classes of a very large group of substances which have as yet no widely accepted name but which might be called biological catalysts. As yet all of the substances clearly belonging in these groups are organic in nature, but it is quite within the bounds of possibility that there exist inorganic substances which exert profound effects in such small amounts that they might almost be called *inorganic vitamins*. Indeed substances of this nature are already known in connection with both plant and animal metabolism.

The problem, then, involved in vitamin research is essentially to discover, by means of carefully controlled experiments, the nature of a substance which is causing a given effect by its presence in extremely small amounts. And since the active substance is organic in nature and its effects are biological, it requires the combined efforts of biologists and chemists to solve the problem. In no other field, perhaps, has there been such an absolute necessity for cooperative efforts among different

## LEE IRVIN SMITH

branches of scientific research as in this field of vitamins and hormones. Since vitamins are obtained largely from foods, the discovery of the existence of a vitamin has usually come about as the result of carefully controlled feeding experiments; indeed our knowledge of vitamins today may be counted largely as one of the triumphs of the modern techniques in the field of nutrition.

The study and elucidation of the nature of a vitamin and of its biological action may be divided roughly into five stages:

1. Recognition of a biological effect not caused by any functional abnormality, which is traced to a lack of something in the "highly purified" diet used, because the biological effect does not result when a well-balanced diet made up of naturally occurring, unprocessed foodstuffs is used.

2. By a process of selective elimination the substance responsible for the biological effect is located in one or more of the dietary components, and a foodstuff is found which is shown to be particularly rich in the unknown substance. Here we have experimental production of the disease state, which can be cured by administration of the active dietary factor. The material found to be rich in the unknown substance is processed in various ways, by extraction, partition, and the like, each of the components—proteins, fats, nonsaponifiable part, etc.—being tested experimentally for its curative action in the disease state. In this way it is discovered that the active factor is concentrated in one or two of the extracts from the foodstuff. And at this point comes the first clue, usually by elimination, as to what sort of a chemical substance the active principle may be. For instance, if it is found that the active principle becomes concentrated in the nonsaponifiable, non-nitrogenous part of the extracts, it follows that the active principle cannot be a fat (ester) nor a protein. Along with this second stage in the study of the vitamin come improvements in the methods of bio-assay; these usually lead, at about this stage, to well-developed, reliable methods which are as simple as is possible under the circumstances.

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

3. At the third stage the entrance of physical and chemical methods begins, and these methods play an ever increasing role as the study develops. Some physical criterion is sought by means of which the presence of the active substance may be determined quickly, without waiting for the results of the time-consuming bio-assays. A very common criterion is the absorption spectrum; often this will show a maximum or minimum at a given point, which correlates with biological activity and so is characteristic of the active substance. The concentrates are worked over and further separated into fractions, one or two of which will show greatly increased biological activity, and the others less than the original concentrates. Soon very active concentrates are obtained, and with these, *chemical* tests are made to determine any active groups that may be present—ketone, aldehyde, hydroxyl, and the like—and to find if possible a reagent which will precipitate the active substance. In any event the active principle, which usually long before this stage has been recognized as a vitamin and given a letter in the vitamin alphabet, is finally isolated in pure form and its composition established by analysis. By the time this stage is reached thousands of bio-assays will have been made, and often much may be known, in a general way, about the substance chemically and biologically.

4. In the fourth stage the methods of organic chemistry are applied to the substance to determine the structure and then to synthesize it in the laboratory. This is usually difficult, because for the most part the vitamins are very complex organic molecules and available only in minute amounts. The chemical degradation of such very small amounts of pure vitamins as are usually available initially in researches of this type is a truly formidable task, and were it not for the modern techniques enabling the chemist to work efficiently with milligrams of material rather than with grams, the task would be quite hopeless. But in this fourth stage the structure is determined and the compound is synthesized. The synthesis always makes available relatively large amounts of the pure vitamin, hence

## LEE IRVIN SMITH

experiments may now be made with potent, standardized, and uniform material at a fraction of the previous cost.

5. The fifth stage, the final one, may be said to be practically endless. The researches at this stage may be subdivided into a) the chemical and b) the clinical.

a) The chemical part of the fifth stage deals with the preparation of analogs of the vitamin in the hope of finding simpler, more readily accessible compounds with high activity. Thousands of new compounds, related structurally to the vitamin, are made; all are tested biologically and any activity is correlated, if possible, with structure. In this way it may be possible to locate the part of the molecule largely responsible for the activity, and having located it, to modify and improve it. If this should be successful a "synthetic vitamin," in the strictest sense of the words, would result. Along with the search for new and simpler compounds of high activity goes, in collaboration with the physiologist, a search for the chemical mechanism whereby the vitamin operates in the body; as examples of active compounds accumulate, certain structural features common to them may enable some possible mechanisms of action—such as oxidation-reduction, etc.—to be excluded while other mechanisms remain still possible.

b) The clinical part of the fifth stage has to do with the use of the vitamin in the treatment of human disease. Nor must one at this stage neglect the veterinary field, for frequently vitamins have found great use in the treatment of domestic animals: this promises to be one of the very important fields of application for vitamin E. From this it is easy to see that vitamins may well be of tremendous importance in the economy of a country and in public health.

In these stages of the study of a vitamin, the first three are the most time consuming. In the beginning of the study of the vitamins a long time elapsed between establishing the existence of a vitamin and isolating it in pure form from natural sources; once the isolation was achieved, the synthesis usually followed fairly quickly. The great amount of time used for the

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

early work was due to a lack of knowledge of procedures and techniques. These had to be developed as the study progressed, and guiding principles had to be established; the methods of nutrition research—controlled diets and the like—also had to be perfected. But with each succeeding study of a vitamin the time intervals between stages have decreased, so that in the most recent vitamin study, that of K, the first four stages were completed in three years. Table 1 shows this trend (1).

TABLE 1.—DISCOVERY OF THE VITAMINS

Vitamin	Existence Established	Isolation	Synthesis	Designation
A .....	1912-15	1928	1937	Vitamin A
B .....	1912-15			
B <sub>1</sub> .....	1912-15	1926	1936	Thiamin
B <sub>2</sub> .....	1933-	1933	1935	Riboflavin
P-P .....	1925-26	1937	1967	Nicotinic acid amide
B <sub>6</sub> .....	1936	1938	1939	Vitamin B <sub>6</sub>
C .....	1919	1932	1933	Ascorbic acid; cevitamic acid
D .....	1922	1927	1927*	Calciferol
E .....	1922	1936	1938	Tocopherol ( $\alpha$ , $\beta$ , $\gamma$ )
K <sub>1</sub> .....	1936	1939	1939	Vitamin K <sub>1</sub>
K <sub>2</sub> .....	1939	1939		Vitamin K <sub>2</sub>

\* Not a true synthesis from simpler substances. As the organic chemist understands the term, vitamin D has not yet been synthesized, although its structure is known with a great deal of certainty.

These stages in the investigation of the nature of a vitamin will now be outlined, using vitamin E (tocopherol) as the example.

*First stage: recognition of the biological effect.*—In 1922 Evans and his collaborators at California (2) described the results of a long series of experiments, which indicated that there was required in animal nutrition a dietary constituent necessary for normal reproduction. Young rats, fed a sufficiently long time on a diet of purified foods with addition of the necessary salts and all of the known vitamins, lost the ability to reproduce. By adding certain vegetable products to the diet the reproductive ability was regained. It followed that there existed, in the added vegetable products, an unknown factor necessary

## LEE IRVIN SMITH

for the normal reproductive ability of rats. Evans designated this as factor-X; later he called it vitamin E. It has also been called the antisterility factor or the reproductive vitamin. Although other vitamins, especially vitamin A, appear to exert an influence on the reproductive ability, this loss is most characteristic of a lack of vitamin E. The existence of vitamin E was at first disputed by several workers, but as the studies progressed it was shown that these workers had used diets not quite free from vitamin E and soon there was general agreement that the factor actually existed. Almost simultaneously with Evans' publications Sure (3) and Mattill (4) published the results of their experiments, which also indicated the existence of the antisterility factor. These results have since been duplicated in many other laboratories.

*Second stage: location of foodstuffs rich in vitamin E, and experimental production of the disease state.*—Further investigation of vitamin E was intensively undertaken by Evans and his associates. Extended series of experiments, involving many thousands of experimental animals, were carried out (5). The results showed that wheat-germ oil was the richest source of vitamin E, but that considerable amounts of the vitamin were also found in cottonseed oil, lettuce oil, rice-germ oil, and other seed-germ oils. The vitamin remains in the unsaponifiable part of the lipid fraction. By processes of partition between different solvents, a sterol-free concentrate was obtained, which was active in single doses of 20 mg. (5).

The characteristic symptoms of lack of vitamin E differ in the sexes. In the female rat (5) normal conception occurs, but this is followed by "resorption sterility." There is the usual pregnancy increase in weight for about 10 days; then the weight decreases, becoming normal at about the twentieth day. No litter is cast. The litter has been resorbed, but the resorption has no effect upon the next oestrus cycle. If a female known to be in this state of resorption sterility is again mated, conception occurs as before. A day or so later the animal is given in the food the substance to be tested. If this is active, the



## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

pregnancy will be terminated by the birth of a litter of living young. The vitamin E activity is usually expressed as milligrams of the substance, fed in a single dose, necessary to cure the sterility and to produce litters in 50 per cent of the animals used (5, 6, 7).

In male animals the characteristic symptoms of lack of the vitamin are associated with the germinal epithelia and the spermatozoa. These degenerate until all sexual power is lost. These changes can be arrested by vitamin E only in the early stages; once the degeneration in the male animal progresses very far, administration of the vitamin is of no use.

Along with these changes in the reproductive organs go other, more obscure degenerative changes elsewhere. Recently Shimotori, Emerson, and Evans (8) have reported on cases of muscular dystrophy caused by lack of vitamin E, and there are growth effects (9) clearly discernible, as well as a characteristic paralysis of the hind quarters (10). Other effects, especially connected with the hypophysis and with the occurrence and growth of tumors, have been reported; but there is not complete agreement, as yet, about the connection between these effects and vitamin E.

*Third stage: isolation.*—At the close of the second stage in the study of vitamin E it was possible to obtain a concentrate from wheat-germ oil which showed activity in doses of 20 mg. These were yellow to red oils which were extremely difficult to concentrate further. By high vacuum distillation Olcott and Mattill (11) were able to obtain a fraction boiling at 200°–250° under 0.05–0.1 mm. pressure, which was active in doses of 5 mg., but the vitamin was damaged in this process by the high temperature necessary for the distillation. Evans subjected the concentrates to partition between petroleum ether and methanol and obtained highly active preparations; Drummond used chromatographic adsorption to achieve the same end. But none of these procedures yielded the vitamin in crystalline form. At each stage in these separations the various fractions were assayed biologically, and also at this time measurement

# LEE IRVIN SMITH

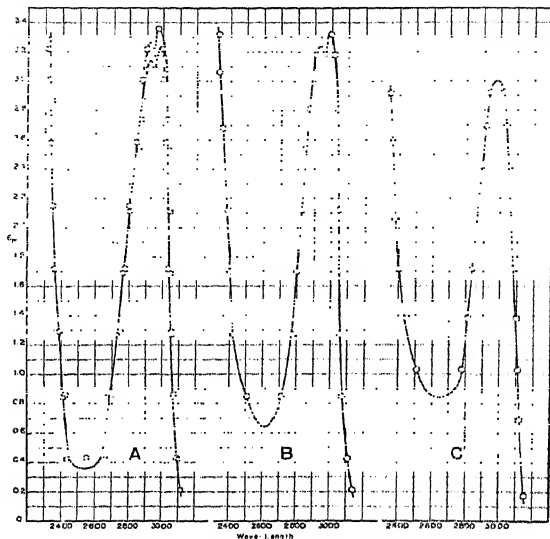
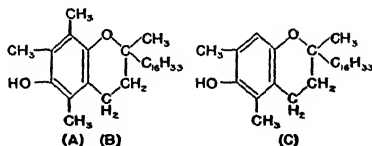


Figure 1. Absorption spectra of tocopherols.



of the ultraviolet absorption spectra of these concentrates was begun. It was found that a parallel existed between the activity and the height of an absorption band at  $2940\text{\AA}$  (12, 13, 54), and this proved to be a reliable guide in following the process of concentration. The curves are given in Figure 1 (54). A is the curve for natural  $\alpha$ -tocopherol, circles and squares representing two different preparations. B is the curve for synthetic *dl*- $\alpha$ -tocopherol, and C is the curve for *m*-xylotocopherol. In Figure 2 are given, for comparison, the curves of three model substances related in structure to the tocopherols; the similari-

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

ties as well as the differences of the chroman and coumaran types are apparent from these curves.

These vitamin E concentrates are readily soluble in all lipid solvents, and only slightly soluble in water. They withstand a temperature of about 200° and are fairly stable in the air when in mass, although when finely divided they are attacked by air and lose their activity. Ultraviolet light quickly destroys all of the activity. The concentrates are fairly stable toward acids, much less so toward alkalis. They are resistant to reduc-

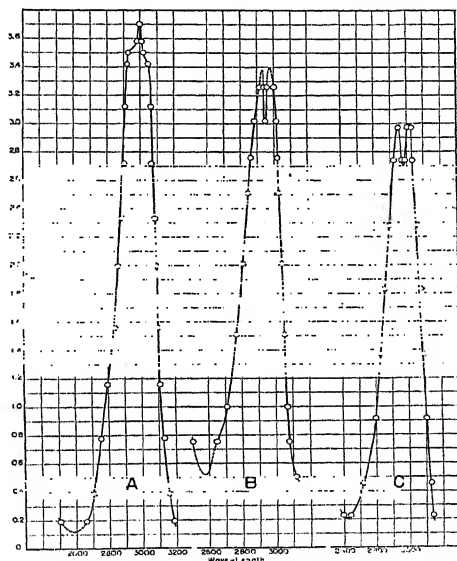
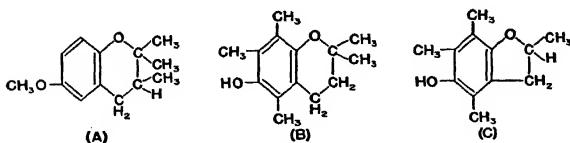


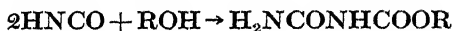
Figure 2. Absorption spectra of simple chromans and coumarans.



## LEE IRVIN SMITH

tion but are quickly attacked by oxidizing agents, even by such mild oxidizing agents as ferric chloride. Acetylchloride and benzoyl chloride react to produce esters, and these esters have practically the same activity as the original material. By comparing the shift in the maxima of the absorption spectra that takes place when phenol is acetylated with that occurring when vitamin E concentrates are acetylated, John (14) was able to deduce that the hydroxyl group in vitamin E was phenolic in nature.

However, esterification of these concentrates by various acids failed to produce solid esters, and it was not until Emerson (15) treated the concentrates with cyanic acid that a solid derivative of the vitamin was obtained. This reaction, characteristic of the hydroxyl group, leads to esters known as allophanates.



By careful purification of the solid obtained in this way from wheat-germ oil concentrates there was obtained first an allophanate melting at  $159^{\circ}$ – $160^{\circ}$  and then a second allophanate melting at  $138^{\circ}$ . These allophanates were hydrolyzed, and each yielded a pale yellow oil. These oils were both highly active, the first in 3 mg. doses, the second in 8 mg. doses. For these individual vitamin E factors Evans coined the name tocopherol; the tocopherols were then designated as  $\alpha$ - and  $\beta$ -tocopherols. From one kilogram of wheat-germ oil, about one gram of  $\alpha$ -tocopherol allophanate may be obtained, although the yield is often much less than this.

$\alpha$ -Tocopherol possesses all of the properties of the highly active concentrates from wheat-germ oil. It shows the same solubility behavior, and the absorption band at  $2940\text{\AA}$  is the same. Analysis shows the composition to be  $\text{C}_{20}\text{H}_{50}\text{O}_2$ . The homogeneity of the preparation was shown by converting it into a solid p-nitrophenyl urethane and a solid p-nitrobenzoate and transforming these into allophanates with the same melting point as that possessed by the original allophanate from the concentrates.

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

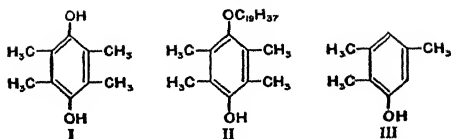
$\beta$ -Tocopherol, obtained in the same way from its allophanate, is likewise an oil. Its properties are almost identical with those of  $\alpha$ -tocopherol, but its composition is  $C_{28}H_{48}O_2$  and so it is a lower homolog of  $\alpha$ -tocopherol. The yield of  $\beta$ -tocopherol from wheat-germ oil is usually much smaller than the yield of  $\alpha$ -tocopherol, but often, from oils of different sources, normal amounts of  $\beta$ -tocopherol can be isolated while almost no  $\alpha$ -tocopherol can be found.

A third allophanate, melting at  $138^\circ$ , has been isolated from cottonseed oil by Emerson, Evans, Olcott, and their associates. This has been named  $\gamma$ -tocopherol allophanate.  $\gamma$ -Tocopherol is likewise an oil, active in 8 mg. doses, and is an isomer of  $\beta$ -tocopherol, having the composition  $C_{28}H_{48}O_2$ .

We have, then, three antisterility factors which are responsible for vitamin E activity. These three tocopherols appear to be the only substances isolated from natural material which certainly possess vitamin E activity, for reports of still other active principles have not been substantiated (90, 20b).

*Fourth stage: determination of the structure, and synthesis.*  
 $\alpha$ -Tocopherol.—As mentioned above,  $\alpha$ -tocopherol possesses the composition  $C_{29}H_{50}O_2$ , which is very close to that of some of the sterols, sitosterol for instance, having the composition  $C_{29}H_{50}O$ . As dehydrogenation with selenium had been of such great value in connection with structure studies in the field of the sterols, it was natural that this method should be applied to  $\alpha$ -tocopherol. McArthur and Watson (16) heated  $\alpha$ -tocopherol with selenium; the result was a yellow sublimate, duroquinone, and a red oil. Somewhat later Fernholz (17) pyrolyzed  $\alpha$ -tocopherol at  $350^\circ$  in the absence of any dehydrogenating agent. There was obtained a good yield of a white crystalline sublimate, identified as durohydroquinone (I), together with a red oil. The simplest assumption which would account for these decomposition products was that  $\alpha$ -tocopherol was a monoether of hydroduroquinone, such as II (in which the group  $C_{19}H_{37}$  contained one saturated ring), for it was known that many alkyl ethers of phenols were cleaved by pyrolysis into

# LEE IRVIN SMITH

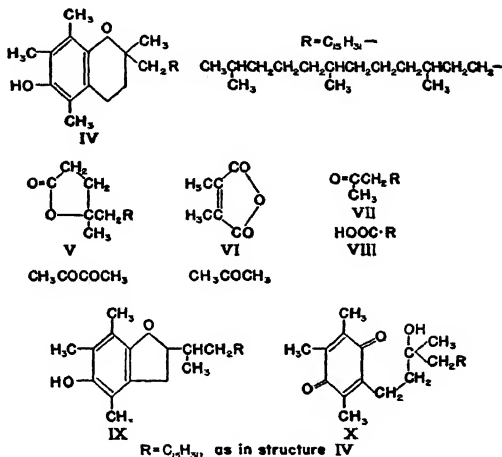


the phenol and an unsaturated hydrocarbon. Accordingly in several laboratories mono-ethers of hydrodurequinone and of other hydroquinones were synthesized. Some of these showed activity when assayed biologically, but these ethers differed markedly from  $\alpha$ -tocopherol in chemical properties, and their ultraviolet absorption spectra were also quite different from that of the vitamin. As a result of these studies it quickly became apparent that  $\alpha$ -tocopherol could not be a simple mono-ether of hydrodurequinone. John, Dietzel, and Günther (14) had also obtained pseudocumenol-6 (iso pseudocumenol) III, by heating  $\alpha$ -tocopherol with hydriodic acid; this result was also difficult to reconcile with the assumption that  $\alpha$ -tocopherol was a simple mono-ether of hydrodurequinone, but it could be reconciled with the assumption that a second ring was condensed with the aromatic nucleus, probably involving an oxygen atom. Bergel, Todd, and Work (18) found that  $\alpha$ -tocopherol, when energetically hydrogenated, absorbed four moles of hydrogen; and they too supposed that an oxide ring was a part of the structure of the vitamin.

The correct structure for  $\alpha$ -tocopherol (IV) was proposed by Fernholz (19) as a result of oxidative degradation, using chromic acid as the oxidizing agent. The products were a  $C_{21}$  lactone (V), dimethylmaleic anhydride (VI), a  $C_{18}$  ketone (VII), a  $C_{16}$  acid (VIII), together with diacetyl and acetone. The hydroxy acid corresponding to the lactone V was transformed into the lactone with extreme ease, indicating that it was a  $\gamma$ -hydroxy acid; moreover, the hydroxyl group of the acid could not be oxidized to a carbonyl group, and was esterified only with difficulty. These facts indicated that the hydroxyl group was tertiary. The  $C_{16}$  acid VIII, when analyzed for  $C-CH_3$  groups, showed three such groups. The structure for

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

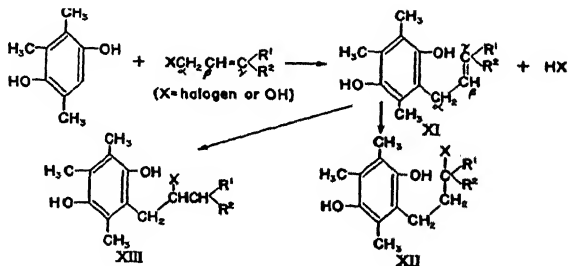
the lactone V can only be written as shown in order to explain the formation from it of a C<sub>18</sub> ketone and a C<sub>16</sub> acid, and when these degradation products are assembled they lead unequivocally to the structure IV, a chroman, for  $\alpha$ -tocopherol. These results do not, of course, lead to the structure shown for the group R, C<sub>15</sub>H<sub>31</sub>. The structure for this group was written on the basis of the C-methyl determination and the experiences gained in other fields of natural products, which frequently contain chains of "isoprene" units joined head to tail.



Karrer (20), although considering both the chroman (IV) and the coumaran (IX) structures for  $\alpha$ -tocopherol, at first preferred the latter. However, John and his associates (21) showed that  $\alpha$ -tocopherol, when oxidized carefully with silver nitrate or ferric chloride, gave a yellow quinone X. This quinone could be reduced to a hydroquinone, the di-p-bromobenzoate of which was quite stable toward chromic oxide, a fact which indicated that the hydroxyl group in X was tertiary. This could only be true if the oxygen ring in  $\alpha$ -tocopherol were a chroman, for the coumaran IX would on oxidation give a hydroxy qui-

none whose hydroxyl group would be secondary and thus susceptible to ready oxidation by chromic oxide. Karrer based his selection of the coumaran formula IX upon the fact that allyl bromide, when condensed with trimethylhydroquinone, does give a coumaran, and when he synthesized  $\alpha$ -tocopherol from phytol bromide, trimethylhydroquinone, and zinc chloride (22) he stated that structure IX was "highly probable."

The synthesis of chromans such as IV is rather simple and easy. The starting materials are hydroquinones (or phenols) having vacant one position in the ring ortho to the hydroxy group (67, 96, 99). These are condensed with allylic halides or alcohols, or with conjugated dienes. Frequently the reaction proceeds so smoothly that neither solvent nor catalyst is required, especially when allylic bromides or chlorides are used. When the alcohols or the dienes are used, it is customary to employ both a catalyst and a solvent. But in any event, because of the great reactivity of the allylic compounds, coupled with the enhanced activity of the aromatic nucleus in polyalkyl benzene derivatives, reactions between the two classes of compounds take place readily, and the products are produced in good yields.

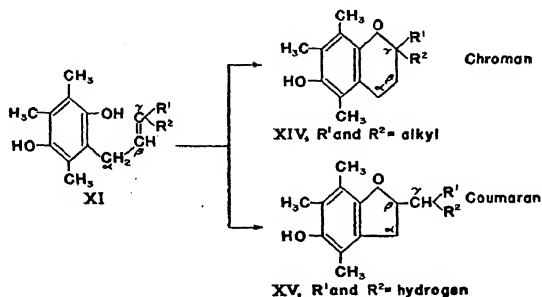


Using the halides or the alcohols, the first step in the reaction appears to be a direct introduction of the allyl group (23, 24) without rearrangement, to give XI. Frequently, when  $X$  is halogen, the  $HX$  addition products XII or XIII of the allylic compounds XI can be isolated. The second step in the reaction,



## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

the ring closure, involves the addition of the hydroxyl group to the double bond in the side chain of XI, in accordance with Markownikoff's rule. Hence, whether a chroman or a coumaran will be formed in this reaction will depend upon the nature of the groups or atoms attached to the  $\gamma$ -carbon atom in the allylic compound. If these groups are both alkyl, the oxygen of

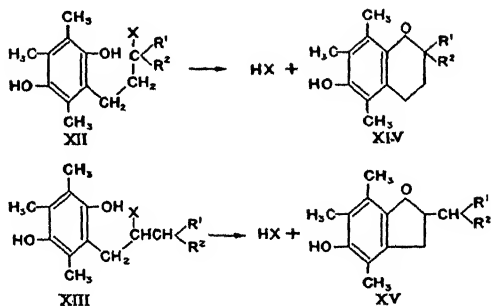


the hydroxyl group will add to the  $\gamma$ -carbon atom, and the product will be a chroman (XIV); while if these groups are both hydrogen, addition will occur in the reverse manner and the product will be a coumaran (XV). When one of the groups is alkyl and the other is hydrogen, the product might be either the chroman or the coumaran or a mixture of the two, although in most of these cases which have been studied so far it is largely the chroman. In a recent paper Karrer, Escher, and Rentschler (102) have made similar generalizations about these ring closures; they have isolated, as condensation products of trimethylhydroquinone and crotyl bromide, *both* the chroman XXXIII and the coumaran XXXIIIa. The structure of the latter was proved by an independent synthesis, using the sequence of reactions shown for the synthesis of XVI, substituting propionyl acetic ester for acetoacetic ester.

The halogen-containing products, XII and XIII, follow the same general rules. These can be readily cyclized to ring compounds, HX being eliminated between the halogen atom and the hydrogen atom of the hydroxyl group. It is to be noted that

LEE IRVIN SMITH

Markownikoff's rule also plays a part in these reactions, for although XII cyclizes to XIV ( $R^1$  and  $R^2$  are alkyl), the addition of HX to the double bond in XI could occur in two ways, and the mode of addition will be governed by the rule. Thus when  $R^1$  and  $R^2$  in XI are alkyl groups, HX will add so as to produce XII; but when  $R^1$  and  $R^2$  are hydrogen atoms, HX will add so as to produce XIII. The ensuing ring closure by elimination of HX would then give the chroman XIV from XII and the coumaran XV from XIII.



The generalities stated above regarding these reactions were carefully checked by means of model experiments upon simple compounds, the structures of which could be proved by independent syntheses. Thus, when allyl bromide or chloride is condensed with trimethylhydroquinone, the product is the coumaran XVI (24, 26), which is also produced by reduction of the coumaron XVII, whose structure had previously been proved (27). When  $\gamma, \gamma$ -dimethyl allyl bromide is used, the ring closure occurs in the reverse direction and a chroman XVIII is produced (24, 26, 28). The structure of this chroman also was proved by an independent synthesis from coumarin derivatives (XIX, XX, and XXI) of known structure (29, 30). This same chroman XVIII has also been synthesized in other ways (31, 32, 33) so there can be no doubt as to its structure; it is also the product of the reaction between trimethylhydro-

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

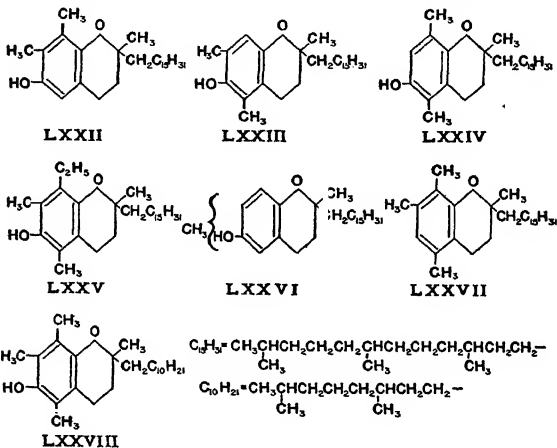
*Specificity of vitamin E.*—Vitamin activity is usually very specific, and even slight changes in the structure of the vitamin molecule are sufficient to reduce the biological activity greatly or to remove it completely. Contrary to the usual experience in this field, vitamin E activity is shown by a great number of compounds widely different in nature and only slightly related in structure to the tocopherols. Over one hundred and thirty individual compounds have been assayed biologically, and a complete list of these has recently been published (65). The list of substances examined includes chromans, chromenes, coumarans, coumarins, coumarons, phenols, quinones, hydroquinones and their esters and ethers. Of the one hundred and thirty or more compounds tested, over forty show vitamin E activity. It is true that with very few exceptions none of these synthetic substances compare with the tocopherols in activity and many of them have to be fed at levels approaching the toxic (50–100 mg.) in order to obtain positive results in the bioassays, yet these results do show quite definitely that vitamin E activity is fairly widespread and by no means confined to a single class of compounds.

$\alpha$ -Tocopherol has been synthesized from trimethylhydroquinone and synthetic phytol (66); this product showed approximately the same activity as natural  $\alpha$ -tocopherol. In all the tocopherol syntheses using phytyl derivatives the product is racemic about the  $\alpha$ -carbon atom of the heterocyclic ring; it is customary to refer to such tocopherol as *dl*- $\alpha$ -tocopherol. This substance has been resolved via the bromo camphor sulfonates (20b, 22), but the activity is unchanged. Hence symmetry or asymmetry about the  $\alpha$ -carbon atom of the heterocyclic ring does not influence the activity of the tocopherol, nor does the optical state of any of the asymmetric carbon atoms in the phytol side chain exert any effect, since the product from synthetic phytol was as active as the natural tocopherol (66, especially 101). In fact, it appears that natural tocopherol is racemic about all three asymmetric centers (101). Synthetic  $\alpha$ -tocopherol and its acetate are nontoxic, and very large doses

LEE IRVIN SMITH

(50g. per kilo body weight to mice) have no ill effects (94). Nor do these substances have any carcinogenic properties (95).

The homologs of  $\alpha$ -tocopherol in which the homology is due to changes in the benzene ring show a decreasing activity as methyl groups are removed, or as methyl groups are replaced by ethyl groups. Thus o-, m-, and p-(*dl*- $\beta$ -) xylotocopherols (LXXII, LXXIII, LXXIV) are active when fed at 5–20 mg. levels (42, 65, 101); the ethyl homolog LXXV is active when fed at 10–16 mg. levels (67, 101). The toluotocopherols (isolated

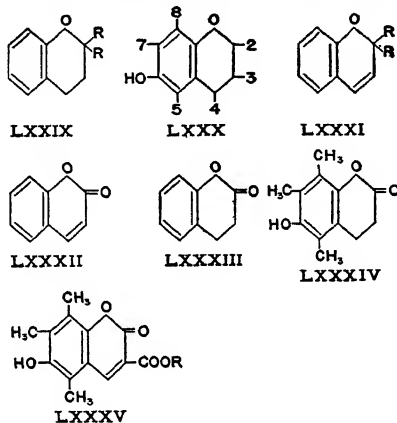


as a mixture; position of the methyl group undetermined) LXXVI are inactive at levels of 40 mg. (42), while 6-desoxy-*dl*- $\alpha$ -tocopherol (LXXVII) is inactive in 100 mg. doses (68). When the phytol side chain in  $\alpha$ -tocopherol is shortened by one isoprene unit, the compound (LXXVIII) is inactive at the 20 mg. level (69). From these results it follows that, starting with  $\alpha$ -tocopherol (activity 3 mg.), any change in the groups in the benzene ring, or in the nature of the long aliphatic side chain, reduces very much the activity of the compound. Further, the hydroxyl group para to the bridge oxygen is necessary for any activity, although it can be masked as any one of several *car-*

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

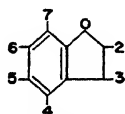
*boxylic* esters without reducing the activity appreciably (52). This hydroxyl group *cannot* be masked as the allophanate, or as an ether, without complete loss of activity.

Turning to the simpler compounds, chromans represented by LXXIX are active when the groups R are hydrogen (68), ethyl (65), or n-butyl (65), but inactive when the groups R are methyl or n-propyl (65). This alternation in activity with groups containing even and odd numbers of carbon atoms is very curious, and it would be interesting to extend the series further. Of the other chromans with the 6-position vacant, 2,

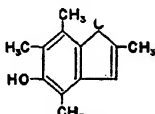


2,3-trimethyl- (65), 2-methyl-4-ethyl- (65), 2,2,5,7-tetramethyl- (65) chromans are inactive, while 2,5,7,8-tetramethyl- (68) chroman is active. A number of chromans with the hydroxyl group in position 6 (LXXX) have been examined. These include 2,5,7,8-tetramethyl- (50, 68, 69), 2,2,5,7,8-pentamethyl- (65), 2,3,5,7,8-pentamethyl- (50), 2,5,7,8-tetramethyl-2-dodecyl- (71), and 2,5,7,8-tetramethyl-2-isohexyl- (65), -6-hydroxychromans, all inactive except the 2,2,5,7,8-pentamethyl compound, which in one test out of three showed a faint activity at the level of 100 mg. (65). The other 6-hydroxychromans studied are all closely related to the tocopherols, and the bio-assays of

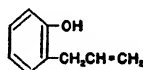
these compounds are discussed in the preceding paragraph. Three chromenes (LXXXI) have been examined (65) — those in which the groups R are methyl, ethyl, and n-butyl. All are inactive. Of the six coumarin derivatives examined (65), coumarin (LXXXII) and dihydrocoumarin (LXXXIII) are inactive; indeed, the former is toxic at the level fed (100 mg.). The dihydrocoumarin LXXXIV, which has the same substituents in the benzene ring as  $\alpha$ -tocopherol, is inactive. Likewise inactive are the substituted coumarins LXXXV when R is hydrogen or isoamyl; but, astonishingly, when R is ethyl the coumarin LXXXV shows a very high activity — it is effective in doses of 20 mg. This compound LXXXV,  $R = C_2H_5$ , is the most active compound known outside of the tocopherols themselves; the activity exceeds that of toluotocopherol LXXVI and is comparable to that of the xylotocopherols LXXII, LXXIII, and LXXIV. This high activity of a compound quite different in structure from the tocopherols is very mysterious, and it becomes all the more inexplicable in view of the inactivity of the two closely related substances obtained when the ethyl group (R) in LXXXV is replaced by hydrogen or isoamyl.



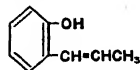
LXXXVI



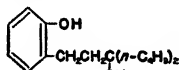
LXXXVII



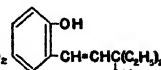
LXXXVIII



LXXXIX



XC



XCI

Several compounds with the heterocyclic ring consisting of five instead of six atoms have been examined. These are for the most part coumarans (LXXXVI), although one coumaron (LXXXVII) has been studied and it is inactive (65). Of the coumarans the unsubstituted molecule (LXXXVI) is inactive (68). 2-Methyl coumaran showed great activity at a level of 50 mg. in one of four assays; the other three assays (25, 50, and

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

100 mg.) were negative (65). 2,2,7-Trimethylcoumaran was also active (65), as was 2,3,4,6,7-pentamethyl-5-hydroxycoumaran (65), while 3-methyl- (65), 2,4,6,7-tetramethyl- (65), 2,4,6,7-tetramethyl-5-hydroxy (65, 68), and 4,6,7-trimethyl-2-*n*-heptadecyl-5-hydroxy- (49) coumarans were all inactive.

Some phenols — mostly containing allylic groups — have been studied (65), since these are possible intermediates in the syntheses of chromans and coumarans from allylic compounds. *o*-Allyl phenol (LXXXVIII) is inactive in 25 mg. doses, but active when the dose is 50 mg. *o*-Propenylphenol (LXXXIX) is inactive. A di-*o*-hexenylphenol (mixture of isomers) is active, as is *p*-amino-*o*-allylphenol. All the other phenols tested are inactive; these included *o*- $\alpha$ -methyl allyl-, *o*-hexenyl-, 2,3,5-trimethyl-6-allyl-, *p*-capryl-, *p*-tert-octyl-, *o*-allyl-*p*-carboxy-, *o*-allyl-*p*-carbethoxy-, and two more complicated phenols XC and XCI.

By far the most extensively investigated compounds, however, are the quinones and hydroquinones, together with esters and ethers of the latter. These compounds were examined in some detail because early in the work on the structure of vitamin E there was some evidence which indicated that possibly the vitamin might be a mono-ether of a methylated hydroquinone. With one exception the para quinones studied are all inactive. These include duroquinone (59, 70), tetraethylquinone (toxic) (65), thymoquinone (toxic) (65), trimethyl ethylquinone (65), 1,4-naphthoquinone (68), 1,2-naphthoquinone (toxic) (65), 2,3-dimethyl-1,4-naphthoquinone (65, 68), 2-methyl-1,4-naphthoquinone (65), 2-hydroxy-1,4-naphthoquinone (65), 2-methoxy-1,4-naphthoquinone (65), anthraquinone (65), and  $\beta$ -methylantraquinone (65). The one exception to the inactive para quinones is  $\alpha$ -tocopherylquinone (X), obtained by mild oxidation of  $\alpha$ -tocopherol, and the results of different workers who have tested this substance do not agree. It has been reported active once (72), but three other assays were negative (57, 68). The red *o*-quinone oxidation product of  $\alpha$ -tocopherol (LXVI,  $R' = CH_3$ ,  $R = C_{15}H_{31}$ ) is inactive in doses of 3 and 6 mg. but active in doses of 12 mg. (65).

# LEE IRVIN SMITH

Of the hydroquinones examined the unsubstituted hydroquinone is inactive (59), as is m-xylohydroquinone (65, 68), but o-xylohydroquinone is active (73), while p-xylohydroquinone is inactive in 50 mg. doses (65) but active in 100 mg. doses (68). Trimethylhydroquinone has been reported as inactive (100 mg.) (65) and also active at this same level (68). Durohydroquinone is active (65, 68) while trimethylethylhydroquinone and trimethyl-5-acetohydroquinone are inactive (68, 92). The only naphthohydroquinone tested is 2,3-dimethyl-5,6,7,8-tetrahydro-1,4-naphthohydroquinone; this compound exhibits good activity (73), as does its mono-*n*-dodecylether (70).

TABLE 3. — ETHERS AND ESTERS OF TRIMETHYLHYDROQUINONE

Derivative	Activity	Reference
Monobenzoate .....	+	68
Bis $\beta$ -iodo propionate.....	—	68
Mono- <i>n</i> -hexyl ether .....	+	73
Mono- <i>n</i> -dodecyl ether .....	+	73
Mono- <i>n</i> -dodecyl ether acetate.....	+	73
Mono-dihydrochaulmoogryl ether .....	+	73
Di- <i>n</i> -dodecyl ether .....	—	73

Table 3 shows the results obtained with a series of esters and ethers of trimethylhydroquinone. Table 4 shows the results obtained with a series of esters and ethers of tetramethylhydroquinone (durohydroquinone).

Three simple phenol ethers, phenylhexenyl ether, phenylcinnamyl ether, and p-carboxyphenylallyl ether have been examined (65); all are inactive. A ketone, 4-(2,5-dimethoxy-3,4,6, trimethylphenyl)-butanone-2, is likewise inactive (65), and finally phytol, alone or in combination with trimethylhydroquinone, shows no activity (65). Hence, even though the synthesis of  $\alpha$ -tocopherol from these compounds in the laboratory is surprisingly easy, the synthesis does not occur in vivo, at least when the substances are fed.

From the results of the bio-assays presented here it is clear that many compounds exhibit some vitamin E activity, and there are one or two fairly simple compounds which show con-



# ORGANIC CHEMISTRY IN VITAMIN RESEARCH

TABLE 4. — ETHERS AND ESTERS OF TETRAMETHYLHYDROQUINONE

Derivative	Activity	Reference
Mono- <i>n</i> -butyl ether .....	+	73
Di- <i>n</i> -butyl ether .....	+	73
Mono- <i>n</i> -hexyl ether .....	+	73
Di- <i>n</i> -hexyl ether .....	+	73
Mono- <i>n</i> -heptyl ether .....	—	73
Di- <i>n</i> -heptyl ether .....	+	73
Mono- <i>n</i> -octyl ether .....	+	73
Di- <i>n</i> -octyl ether .....	+	73
Mono-cetyl ether .....	+ (twice) — (once)	59, 74
Mono-dodecyl ether propionate.....	—	73
Mono-dodecyl ether palmitate.....	+	73
Mono-dodecyl ether .....	+	59, 73, 74
Di-dodecyl ether .....	—	73
Mono-hydrophytyl ether .....	+	73
Mono- <i>n</i> -octadecyl ether .....	—	59, 74
Mono- <i>n</i> -nonadecyl-2 ether .....	+	59, 74
Mono-2-methyloctadecyl ether .....	—	59, 74
Mono- <i>n</i> -nonadecyl ether .....	+	68
Di- <i>n</i> -nonadecyl ether .....	—	68
Mono-3-methyl-5-(1',1',3'-trimethyl-2'-cyclohexyl)-pentyl-1 ether .....	—	68
Mono-dihydrochaulmoogryl ether .....	+	73
Mono-benzyl ether .....	+	73
Di-benzyl ether .....	+	73

siderable activity. These results, until recently at variance with all other results in the vitamin field, appear now to be paralleled in the field of the K vitamins, where a number of substances aside from the vitamins themselves possess great potency. Notable among these are certain 1,4-naphthoquinones, especially 2-methyl-1,4-naphthoquinone. None of these naphthoquinones possess any vitamin E activity, but the methyl naphthoquinone has almost as much antihaemorrhagic activity as vitamin K<sub>1</sub> itself.

There is no adequate theory at present to account for vitamin E activity in terms of organic structure. One theory has been advanced (50, 57, 65) independently from two laboratories, but there remain objections to the theory which will have to be overcome before it can be accepted (50, 57, 65, 75).

*Uses and importance of vitamin E.*—Vitamin E appears to

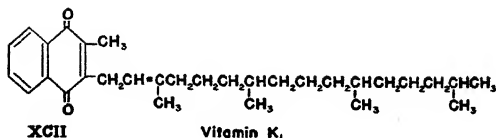
## LEE IRVIN SMITH

be a most promising substance to be used in the treatment of habitual abortion in women, and for similar use in the veterinary field. Some rather startling successes have been reported when the vitamin has been used in these cases. There is good evidence also that the young of both sexes need vitamin E for normal growth, and that there is some connection between the amount of vitamin E and the functioning of the thyroid gland as well as that of the hypophysis. That muscular dystrophy can result from a lack of vitamin E appears to be well established (8). The writer is not competent to discuss this field, and since reviews covering the use of vitamin E in medicine have recently been published (76-85, especially 83, 85) only the most general statements have been made here. Most of the studies so far have been carried out using wheat-germ oil concentrates, but now that the synthetic vitamin is available in pure form a standard preparation of known potency is available; it is to be hoped that the clinical work will proceed rapidly so that the usefulness as well as the limits of vitamin E therapy may soon be known.

*Vitamin K.*—Before closing this review a word about vitamin K may not be out of place, for it is unique in being the first of all the vitamins whose chemistry has been aided and simplified by the chemical knowledge gained in the study of any other vitamin. Until the recent work on the structure and synthesis of vitamins E and K no vitamin had ever been found which was in any way related chemically to any other vitamin—each vitamin belonged to an entirely different class of chemical compounds; indeed, vitamins A and D might almost be said to have represented, at the time they were isolated, new classes of organic compounds. Vitamins E and K, however, are closely related in their chemistry, and the knowledge and experience gained in studying one of these vitamins has been of great value in the study of the other. There are two K vitamins, K<sub>1</sub> (XCII) and K<sub>2</sub>, both of which show a powerful antihæmorrhagic activity although K<sub>1</sub> surpasses K<sub>2</sub> in this respect.

Vitamin K<sub>1</sub> has been synthesized in two laboratories (87,

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH



88). However, the methods used for the synthesis of vitamin E had to be modified considerably in order to avoid ring closures which would lead to compounds of the tocopherol type. Unfortunately it has not been found possible to convert compounds of the tocopherol types into compounds of the vitamin K types, although the reverse transformation appears readily possible. Vitamins E and K also have in common the fact that the specificity is not limited to the vitamins themselves—indeed, 2-methyl-1,4-naphthoquinone appears to be fully as active as vitamin K. Curiously enough, no compounds have been found so far which show both kinds of activity: if a substance exhibits vitamin E activity it does not show any vitamin K activity, and vice versa.

## References

References 76-85 inclusive are review articles; the whole issue of *Angew. Chem.* 52:413-32 (No. 24, July 17, 1939) was devoted to the review articles of John (82) and Grandel (83) and to a verbatim report of the Vitamin E Conference held in London April 22, 1939 (84, 103).

1. Table adapted from Mason, K. E. 1939. Vitamins and hormones. Science in Progress. Yale Univ. Press. p. 147.
2. a) Evans and Bishop. 1922. Science 56:650; b) 1923. Am. Jour. Physiol. 63:396; c) 1922, 1923. Jour. Metabol. Res. I:319, 335; III:202, 233; d) 1922. Am. Jour. Physiol. 63:396; e) 1922. Jour. Am. Med. Assn. 81:889; f) 1924. Anat. Rec. 27:203; g) 1922. Science 56:650; h) Evans. 1932. Jour. Am. Med. Assn. 99:470.
3. Sure. a) 1924. Science 59:19; b) 1924. Jour. Biol. Chem. 58:693; c) 1925. Jour. Biol. Chem. 62:371.
4. a) Mattill and Conklin. 1920. Jour. Biol. Chem. 44:137; b) Mattill and Stone. 1923. Jour. Biol. Chem. 55:443; c) Mattill, Carmen, and Clayton. 1924. Jour. Biol. Chem. 61:729.
5. Evans and Burr. 1927. Memoirs of the University of California. No. 8.
6. Palmer. 1937. Indust. Eng. Chem. Anal. Ed. 9:427.
7. Bacharach et al. a) 1938. Nature 142:35; b) 1937. Biochem. Jour. 31:

# LEE IRVIN SMITH

- 2287; c) 1938. *Biochem. Jour.* 32:1298; d) 1938. *Jour. Soc. Chem. Indust.* 57:600.
8. Shimotori, Emerson, and Evans. 1939. *Science* 90:89.
9. a) Evans. 1928. *Jour. Nutr.* 1:1, 23; b) Emerson and Evans. 1937. *Jour. Nutr.* 14:169.
10. Evans and Burr. 1928. *Jour. Biol. Chem.* 76:273.
11. Olcott and Mattill. 1934. *Jour. Biol. Chem.* 104:423.
12. Bowden and Moore. 1933. *Nature* 131:512; 132:204; Martin, Moore, and Schmidt. 1934. *Nature* 134:214.
13. Drummond and Hoover. 1937. *Biochem. Jour.* 31:1852.
14. John, Dietzel, and Günther. 1938. *Zeitschr. f. physiol. Chem.* 252:208.
15. Emerson and Emerson. 1936. *Jour. Biol. Chem.* 113:319.
16. McArthur and Watson. 1937. *Science* 86:35.
17. Fernholz. 1937. *Jour. Am. Chem. Soc.* 59:1154.
18. Bergel, Todd, and Work. 1938. *Jour. Chem. Soc.* 1938:253.
19. Fernholz. 1938. *Jour. Am. Chem. Soc.* 60:700.
20. a) Karrer, Salomon, and Fritzsche. 1938. *Helv. Chim. Acta* 21:309; b) Karrer, Fritzsche, Ringier, and Salomon. 1938. *Helv. Chim. Acta* 21:820.
21. John, Dietzel, and Günther. 1938. *Zeitschr. f. physiol. Chem.* 252:222.
22. Karrer, Fritzsche, Ringier, and Salomon. 1938. *Helv. Chim. Acta* 21:520.
23. Smith, Ungnade, Stevens, and Christman. 1939. *Jour. Am. Chem. Soc.* 61:2615.
24. Smith, Ungnade, Hoehn, and Wawzonek. 1939. *Jour. Org. Chem.* 4:305.
25. ———. 1939. *Jour. Org. Chem.* 4:311.
26. Smith, Ungnade, and Prichard. 1938. *Science* 88:37.
27. Smith and MacMullen. 1936. *Jour. Am. Chem. Soc.* 58:629.
28. Smith, Ungnade, and Prichard. 1939. *Jour. Org. Chem.* 4:358.
29. Smith and Dobrovolny. 1926. *Jour. Am. Chem. Soc.* 48:1693.
30. Smith and Denyes. 1936. *Jour. Am. Chem. Soc.* 58:304.
31. John and Günther. 1939. *Ber.* 72:1649.
32. John and Schmeil. 1939. *Ber.* 72:1653.
33. Smith, Ungnade, Opie, Prichard, Carlin, and Kaiser. 1939. *Jour. Org. Chem.* 4:323.
34. Claisen and Eisleb. 1913. *Ann.* 401:26; Claisen. 1925. *Ann.* 442:210; Claisen and Tietze. 1925. *Ber.* 58:279; Claisen. 1924. *Ger. Pat.* 394797.
35. Smith and Ungnade. 1939. *Jour. Org. Chem.* 4:298.
36. Bergel, Copping, Jacob, Todd, and Work. a) 1938. *Jour. Chem. Soc.* 1938:1382; b) 1938. *Nature* 142:36.
37. Karrer and Fritzsche. 1938. *Helv. Chim. Acta* 21:1234.
38. John. 1937. *Zeitschr. f. physiol. Chem.* 250:11.
39. Bergel, Todd, and Work. 1938. *Jour. Chem. Soc.* 1938:253.
40. John, Dietzel, and Günther. 1938. *Zeitschr. f. physiol. Chem.* 252:208.
41. Emerson. 1938. *Jour. Am. Chem. Soc.* 60:1741.
42. Karrer and Fritzsche. 1939. *Helv. Chim. Acta* 22:260.
43. Karrer, Fritzsche, and Escher. 1939. *Helv. Chim. Acta* 22:661.
44. Jacob, Steiger, Todd, and Work. 1939. *Jour. Chem. Soc.* 1939:542.
45. Emerson, Emerson, and Evans. 1939. *Science* 89:193.

## ORGANIC CHEMISTRY IN VITAMIN RESEARCH

46. Smith, Hoehn, and Ungnade. 1939. *Jour. Org. Chem.* 4:351.
47. Karrer and Fritzsche. 1939. *Helv. Chim. Acta* 22:657.
48. Bridge, Crocker, Cubin, and Robertson. 1937. *Jour. Chem. Soc.* 1937: 1530.
49. Bergel, Jacob, Todd, and Work. 1938. *Jour. Chem. Soc.* 1938:1375.
50. John, Günther, and Schmeil. 1938. *Ber.* 71:2637.
51. Smith, Ungnade, Austin, Prichard, and Opie. 1939. *Jour. Org. Chem.* 4:334.
52. Demole, Isler, Ringier, Salomon, and Karrer. 1939. *Helv. Chim. Acta* 22:65.
53. Olcott and Emerson. 1937. *Jour. Am. Chem. Soc.* 59:1008.
54. Webb, Smith, Bastedo, Ungnade, Prichard, Hoehn, Wawzonek, Opie, and Austin. 1939. *Jour. Org. Chem.* 4:389.
55. Isler. 1938. *Helv. Chim. Acta* 21:1756.
56. Smith, Irwin, and Ungnade. a) 1939. *Jour. Am. Chem. Soc.* 61:2424; b) 1939. *Science* 90:334.
57. John, Dietzel, and Emte. 1939. *Zeitschr. f. physiol. Chem.* 257:173.
58. Karrer, Fritzsche, and Escher. 1939. *Helv. Chim. Acta* 22:661.
59. Evans, Emerson, and Emerson. 1938. *Science* 88:38.
60. Furter and Meyer. 1939. *Helv. Chim. Acta* 22:240.
61. Karrer and Keller. a) 1938. *Helv. Chim. Acta* 21:1161; b) 1939. *Helv. Chim. Acta* 22:253, 617.
62. Karrer, Escher, Fritzsche, Keller, Ringier, and Salomon. 1938. *Helv. Chim. Acta* 21:939.
63. Emmerie and Engel. 1938. *Rec. Rav. Chim.* 57:1351.
64. Ungnade and Smith. 1939. *Jour. Org. Chem.* 4:397.
65. Evans, Emerson, Emerson, Smith, Ungnade, Prichard, Austin, Hoehn, Opie, and Wawzonek. 1939. *Jour. Org. Chem.* 4:376.
66. Karrer and Ringier. 1939. *Helv. Chim. Acta* 22:610.
67. Karrer and Hoffmann. 1939. *Helv. Chim. Acta* 22:654.
68. von Werder, Moll, and Jung. 1939. *Zeitschr. f. physiol. Chem.* 257:129.
69. Karrer and Jensen. 1938. *Helv. Chim. Acta* 21:1622.
70. John and Günther. 1938. *Zeitschr. f. physiol. Chem.* 254:51.
71. John. 1939. *Angew. Chem.* 52:418.
72. Evans, Emerson, and Emerson. 1936. *Jour. Biol. Chem.* 113:322.
73. von Werder and Moll. 1938. *Zeitschr. f. physiol. Chem.* 254:39.
74. Fernholz and Finkelstein. 1938. *Jour. Am. Chem. Soc.* 60:2402.
75. Karrer. 1939. *Helv. Chim. Acta* 22:349.
76. Evans. 1932. *Jour. Am. Med. Assn.* 99:469.
77. Mattill. a) 1938. *Jour. Am. Med. Assn.* 110:1830; b) 1939. Chapter 30 in *The Vitamins* (a symposium under the auspices of the American Medical Association, published by the association).
78. John. 1938. *Naturwiss.* 26:449.
79. Karrer and Schweiz. 1938. *Med. Wochenschr.* 68:954.
80. Bergel. 1939. *Chem. Prod. Chem. News* 2:39.
81. Karrer. 1939. *Helv. Chim. Acta* 22:334.
82. John. 1939. *Angew. Chem.* 52:413.

# LEE IRVIN SMITH

83. Grandel. 1939. *Angew. Chem.* 52:420.
84. Vitamin E Conference in London, April 22, 1939. Reported in full in *Angew. Chem.* 52:427 (1939).
85. John. 1939. *Ergebn. d. Physiol. biol. Chem. u. expl. Pharm.* 42:1.
86. John and Emte. 1939. *Zeitschr. f. physiol. Chem.* 261:24.
87. Binkley, Cheney, Holcomb, McKee, Thayer, MacCorquodale, and Doisy. 1939. *Jour. Am. Chem. Soc.* 61:2558.
88. a) Fieser, Campbell, Fry, and Gates. 1939. *Jour. Am. Chem. Soc.* 61:2559; b) Fieser. 1939. *Jour. Am. Chem. Soc.* 61:2559, 2561.
89. Jacob, Steiger, Todd, and Work. 1939. *Jour. Chem. Soc.* 1939:542.
90. John. 1938. *Zeitschr. f. physiol. Chem.* 252:201.
91. Karrer, Salomon, and Fritzsche. 1938. *Angew. Chem.* 51:309.
92. von Werder and Jung. 1938. *Ber.* 71:2650.
93. Karrer and Escher. 1939. *Helv. Chim. Acta* 22:264.
94. Demole. 1939. *Zeitschr. Vitaminforsch.* 8:338.
95. ———. 1939. *Zeitschr. Vitaminforsch.* 8:341.
96. Smith, Opie, Wawzonek, and Prichard. 1939. *Jour. Org. Chem.* 4:318.
97. Smith, Ungnade, Lauer, and Leekley. 1939. *Jour. Am. Chem. Soc.* 61:3079.
98. Smith and Prichard. 1939. *Jour. Org. Chem.* 4:342.
99. Emerson and Smith. 1940. *Jour. Am. Chem. Soc.* 62:141.
100. Smith, Ungnade, and Irwin. 1940. *Jour. Am. Chem. Soc.* 62:142.
101. Karrer, Koenig, Ringier, and Salomon. 1939. *Helv. Chim. Acta* 22:1139.
102. Karrer, Escher, and Rentschler. 1939. *Helv. Chim. Acta* 22:1287.
103. Vitamin E, a symposium held under the auspices of The Food Group (Nutrition Panel) of the Society of Chemical Industry, on Saturday, April 22, 1939. Published by the society, Clifton House, Euston Road, London, N.W.1. 88 pp. 13.5 x 21 cm. 5 shillings. For a review of this book see: 1940. *Jour. Am. Chem. Soc.* 62:458.

# ON THE NECESSITY OF FATS IN THE DIET

BY

George O. Burr

*Professor of Botany and Physiological  
Chemistry, University of Minnesota*

SINCE there is little evidence that any special natural form of lipid molecule is required in the diet the problem centers largely on the fatty acids themselves. Extraneous components of an "ether extract" such as pigments, vitamins, and essential oils will not be considered here. Numerous excellent reviews of the physiology of fats, phospholipids, and sterols have appeared in recent years, and an attempt will be made to avoid duplication of their points of view. Necessarily many of the same references will be given (1-5).

*Historical review.*—Since the publications of Liebig in 1843 it has been recognized by physiologists that animals can produce fat from other foods, especially protein and carbohydrate. By 1849 Regnault and Reiset had shown that the R.Q. depended upon the nature of the food given to animals. By the use of this technique as well as by direct analysis of food and tissues it was well established before the end of the nineteenth century that both plants and animals store fat produced from carbohydrates. It was further established that animals synthesize fatty acids not found in common plant seeds, notably the higher acids of fish-liver oils. The discovery of arachidonic acid by Hartley in 1909 showed the presence of an acid peculiar to animals, found in all tissues examined. The further announcement by Leathes (6) of the desaturation of fats in the liver was added evidence that animals were independent of plants for their supply of both physiologically active lipids (element constant of Terroine, 7) and fat reserves.

*Note.* Assistance in the preparation of these materials was furnished by the personnel of Work Projects Administration Official Project No. 665-71-3-69 Sub-project No. 232.

## GEORGE O. BURR

In spite of the general acceptance of this view there were those who maintained an interest in dietary fats by presenting evidence that fats were not wholly replacable by carbohydrate. Foremost among these was Stepp (8), who between the years 1909 and 1911 showed that mice soon died when fed a diet extracted with alcohol and ether. They were saved by the addition of alcohol-ether-soluble extracts, whereas known fats like butter, lecithin, cholesterol, tripalmitin, triolein, and tristearin failed. At just this period the use of isolated food substances in nutrition studies was being adopted. Stepp's significant results led to more work with fat-poor diets by Osborne and Mendel (9), who stated that "with respect to the actual requirement of fat on the part of the healthy organism there is at present almost no definite information available." Their studies at this time failed to establish the need for simple fats, although their rats were subnormal in weight. The continued use of fat-poor diets led to the announcement of vitamin A by McCollum and by Osborne and Mendel in 1913. With this discovery it became evident that earlier failures obtained with low-fat diets may have been due to vitamin deficiency. The center of interest shifted from fats per se to the fat-soluble vitamin, and little more experimental work was done until 1918 when Aron (10) fed rats a diet of protein, wheat bran, and wheat starch. The rats reached a maximum weight of 120 grams and died before they were 160 days old. He postulated a fat minimum as well as a protein minimum. In view of our present knowledge of fat-deficient growth curves and ages of survival, it seems certain that Aron was dealing with a vitamin deficiency. However, this paper was important in that it aroused a new interest in the role of fat in the diet.

During the World War numerous cases of human edema and dropsy were reported in Europe. The observers attributed this syndrome to: (1) lack of fats; (2) lack of vitamin A; (3) general underfeeding (less than 1,400 Cal. per day); (4) lack of both fats and protein; (5) lack of protein; (6) increased work with underfeeding. Salkowski (11) argued that there was little likeli-



## FATS IN THE DIET

hood of fat deficiency because of the great power of synthesis shown by animals and man. Von Gröer (12) cited his experiment with infants as proof that fats and carbohydrates are wholly interchangeable. He stated that the fat minimum is so low that it can be practically disregarded, in spite of the fact that the infants failed to thrive on skim milk after six months. From a series of feeding experiments Maignon concluded that fats are an important part of the diet.

One of the more important results of this controversy was the stimulation of new work giving more attention to the vitamins. McCollum and Davis (13) had shown in 1914 that vitamin A withstood saponification, and this fact was made use of by Drummond and Coward (14) in their study. However, the construction of a diet devoid of true fats but adequate in all other respects was not simple. Small differences in the many mixtures used during the past eighteen years have led to contradictory results which even at the present time have not been entirely eliminated.

In 1920 Kohman (15) fed to rats diets of carrots supplemented by starch, fats, casein, and salts. Severe stunting accompanied by edema occurred when casein was omitted. Fat had no effect. She concluded that "war edema" was more likely due to general underfeeding or protein deficiency than to fat deficiency. At the same time Osborne and Mendel (16) reported new experiments with diets of meat residue, starch, and salts, supplemented by yeast and alfalfa. The rats were "normal," and they concluded that "if true fats are essential for nutrition during growth the minimum necessary must be exceedingly small." In the light of more recent work it is known that their diets carried sufficient fat to mask major deficiencies.

In this same year Krogh and Lindhard (17) produced a new line of evidence in support of a fat minimum. From measurements of the efficiency of work by subjects on an ergometer they advanced the hypothesis that neither fats alone nor carbohydrates alone are suitable for the energy supply of the body. The highest efficiency is reached when the R. Q. is between 0.8

## GEORGE O. BURR

and 0.9. Drummond and Coward (*loc. cit.*) were the first to saponify their cod-liver oil in order to furnish adequate vitamin A without fatty acids. They also included lemon juice and yeast extract for other vitamins. The casein and starch were extracted with hot alcohol and ether. The fat was purified tristearin. Their curves show that 15 per cent of tristearin has no beneficial effect as a supplement to the basal diet. Their work supported the view of Osborne and Mendel and of Hindhede (18) that fats are dispensable if the vitamins are ensured. However, no more unfortunate choice of fat could have been made, since tristearin is totally saturated and poorly digestible.

*Digestibility of fats.*—Although most of the work with very low fat diets done prior to 1920 must be discounted because of the probable deficiency of fat-soluble vitamins, there is no reason to doubt the earlier views that fat is an excellent food, subject to certain limitations of absorption and utilization. These are important considerations when diets carry as much as 31 per cent of the calories as fat, as in the case of the American soldier (Murlin, 19); or 50 per cent, as with the nursing infant. Atwater considered a diet suitable if fat furnished one-third of the total calories. There is over 98 per cent utilization by rats when as much as 86 per cent of the total energy is furnished by fat (20).

It appears that fats are absorbed only after being split into glycerol and fatty acids (21). Therefore physical state is important insofar as it affects contact with enzymes. Absorption begins in the stomach if an emulsified fat is fed, whereas the ordinary fat is not attacked by lipases until it is in the intestine, where bile plays an important role. The free fatty acids are absorbed into the intestinal mucosa, where they may be synthesized into phospholipids and passed into the circulation stream as neutral fat, chiefly through lymph vessels. Sinclair (22) represented this in the following way:

fatty acids  $\longleftrightarrow$  phospholipid  $\longleftrightarrow$  neutral fat

More recently it has been shown that the phospholipids themselves may enter the lymph (Sullman and Willbrandt, 23).

## FATS IN THE DIET

New evidence of phosphorylation of fats in the intestinal mucosa appears periodically (Verzár and Laszt, 24). Recently Miller, Barnes, Kass, and Burr (25) have shown by the use of spectroscopically active fats that as much as 15 per cent of the phospholipids in the intestinal mucosa may be made of fat fed six hours earlier. A new review of fat transport in the animal body has just been published by Bloor (26).

TABLE 1. — THE MEAN PERCENTAGES OF FAT ABSORBED BY RATS FED 1.5 CC. OF FAT \*

Kind	Absorption Time				
	2 hours	4 hours	6 hours	8 hours	12 hours
Lard .....	24.1 ± 0.8	57.0 ± 1.5	67.5 ± 1.5	92.3 ± 0.9	97.8 ± 0.4
Corn oil .....	23.9 ± 0.8	53.3 ± 0.9	71.4 ± 2.1	94.4 ± 0.7	97.9 ± 0.3
Shortening B .....	27.1 ± 1.8	52.8 ± 2.4	71.1 ± 1.5	85.6 ± 1.2	99.6 ± 0.1
Shortening A .....	26.6 ± 1.5	53.8 ± 1.6	68.5 ± 1.7	86.0 ± 1.3	98.6 ± 0.3
Butter fat .....	36.2 ± 1.6	60.3 ± 1.2	77.2 ± 2.0	91.2 ± 1.1	97.4 ± 0.4
Butter oil .....	37.4 ± 2.3	71.0 ± 1.2	86.4 ± 1.7	95.6 ± 1.0	.....
Halibut-liver oil...	39.4 ± 1.6	70.2 ± 2.0	78.1 ± 1.3	85.4 ± 0.9	.....
Cod-liver oil .....	40.8 ± 1.4	67.7 ± 1.9	79.7 ± 1.5	89.2 ± 0.7	98.2 ± 0.4

\*All of the animals were adult male rats. Each figure is the mean of ten or more individual experiments except in the case of the 12-hour group fed cod-liver oil, which figure is a mean of only five experiments.

Digestibility of nearly all fats is essentially the same, providing they have not too high a melting point. The extensive studies by Langworthy (27) with human subjects show over 95 per cent absorption of fats, with a melting point not over 42°. With melting points above 50° the absorption is materially less complete. On complete hydrogenation fats become so hard as to be injurious to the animal. Tofte (28) has recently shown that digestion by pancreatic lipase decreases with increasing melting point of hydrogenated oils.

By a new technique the rate of absorption of several oils was recently measured by Steenbock, Irwin, and Weber (29). Their results are summarized in Table 1. This work shows that early rate of absorption is not closely correlated with any one property like unsaturation or length of fatty acid chain.

Reviews of this subject have been written by Verzár (30), by Holt and coworkers (31), and by Verzár and McDougall (32).

## GEORGE O. BURR

The latter point out that no one factor, melting point for example, accounts wholly for the differences in absorption of fats.

Digestibility measurements indicate that large amounts of stearic and palmitic acid can be digested as a part of mixed glycerides. However, when fed alone as free acids, sodium salts, or pure glycerides, their absorption is very incomplete — 40 to 70 per cent (33, 34).

It would appear from Barbour's work (35) that the arachidic acid of peanut oil is almost wholly excreted in the feces, along with appreciable amounts of the stearic and palmitic of mixed glycerides.

*Fatty acid utilization as affected by combination.*—Fatty acids may be fed as phosphatids; glycerides; methyl, ethyl, or glycol esters; sodium or ammonium salts; or the free acids. Presumably the fatty acids are absorbed as such and the same results should be obtained from all the above forms. But this is not the case, owing either to poor palatability, abnormal absorption rate, or the toxic effects of the accompanying radicle. Some free fatty acids are not good food because they irritate the mouth and are unpalatable. There is evidence that lipids may be absorbed without previous hydrolysis (32), and if this is the case the form of combination would be important. Verzář and Laszt found increased absorption from an intestinal loop of a mixture of oleic acid and bile salts upon the addition of glycine and phosphate.

The extensive comparative study by Ozaki (*loc. cit.*) of the nutritive value of triglycerides, ethyl esters, and sodium soaps indicates that rats will grow well on diets containing 10 per cent of any of the three. The sodium soaps seem to be somewhat inferior to the glycerides and ethyl esters. The poor growth on stearic and palmitic is due to poor digestibility of all forms (see Figure 1).

Lepkovsky, Ouer, and Evans (36) fed the mixed acids from lard as free acids and as esters of methyl and ethyl alcohol, glycerol, ethylene glycol, and propylene glycols. Even when making up 60 per cent of the diet no actual injury resulted from

## FATS IN THE DIET

any of these forms except the ethylene glycol ester, which was toxic and caused kidney lesions. But the glycerol esters gave the best growth at this very high level.

It is known also that the sodium soaps are much more ketogenic than the ammonium salts, a fact which suggests an important effect of the sodium ion on the general metabolism of the animal.

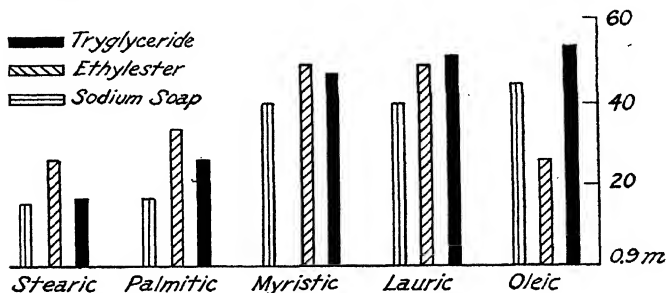


Figure 1. Comparative growth of rats on diets containing 10 per cent of one of the above fatty acids as the soap, ethyl ester, or glyceride. (From Ozaki.)

*Fatty acid utilization as affected by molecular weight or chain length.*—It has already been indicated that as the melting points of the fatty acids rise well above body temperature, digestibilities fall. Appreciable amounts of stearic and practically all the arachidic acid appear in the feces. However, after absorption from the intestine there are other marked differences in utilization, which result from the number of carbons in the chain.

Natural fats are composed almost entirely of fatty acids of even carbon numbers. Hence most of the extensive feeding work has been done with these acids. As early as 1882 Lebedev showed that neither butter nor tributyrin in the diet increased the volatile acids of the body fat of the dog. Animal fats are almost devoid of all acids below myristic and only in recent years has the nutritive value of the short chain acids been determined. Since coconut oil and butter are among the best

## GEORGE O. BURR

food fats it is obvious that the shorter acids are readily utilized when taken as a part of a mixed fat.

Triacetin is an excellent food and at a 10 per cent level is equal to tripalmitin (see Figure 2). Rats grow well on a diet containing 55 per cent of it (37).

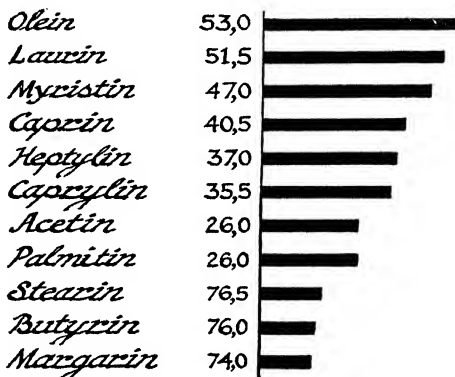


Figure 2. Relative growth of rats fed diets containing 10 per cent of the above synthetic glycerides. (From Ozaki.)

In contrast, tributyrin is so bitter that animals will not eat much of it. When introduced by stomach tube or by injection, toxic symptoms result from large amounts (38). Rats will gain some weight on a diet containing 10 per cent butyryn (see Figure 2), but at very high levels they quickly starve (37).

Caproic, caprylic, capric, lauric, myristic, palmitic, and stearic acids (6 to 18 carbons) have been fed by Ozaki (33), Eckstein (39, 40), Powell (41, 42), Cox (37), and others as ethyl and glyceryl esters and as salts. At a 10 per cent level the glycerides of short acids are all superior to stearin and palmitin. When making up 55 per cent of the diet the ethyl esters of all except myristic cause the death of young rats. The lower acids appear to be toxic at this very high level, while palmitic and stearic are poorly absorbed and cause intestinal obstruction.

## FATS IN THE DIET

All of the acids are readily burned when taken in moderate amounts, but it is not possible to deposit any of the 4, 6, or 8 carbon acids in adipose tissue. Instead they seem to be utilized by the rat in building the higher saturated acids (39, 40, 41, 42). There is a sharp break at this point, and as much as 15 and 25 per cent respectively of the 10 and 12 carbon acids have been deposited by heavy feeding.

So little is known about the exact mechanism of fat burning that no positive statement can be made; however, it seems that all the even carbon acids burn in the same way. Under proper conditions butyric and higher acids give rise to ketone bodies. Several bits of evidence indicate that more than one molecule of beta-oxybutyric acid comes from each of the higher fatty acids, a result that requires the breaking of the chain into 4 carbon units rather than the stepwise burning of 2 carbons at a time by beta-oxidation.

In 1928 Quick (43) extended the work of Knoop and of Dakin, making an 80 per cent recovery of the fed phenylpropionic and phenylbutyric acids in twenty-four hours. He concluded that fatty acids are broken down exclusively by beta-oxidation in the dog. However, this kind of experiment is open to the criticism that methyl burning is blocked by the phenyl groups.

That methyl burning (widely known as omega-oxidation) can occur is proved by the appearance of dicarboxylic acids in the urine (Verkade, 44). Other points of oxidative attack are also postulated and the relative amounts of each probably change with conditions. One point seems to be clear—that in most of the burning an even number of carbons is lost at a time. This results in a striking difference between even and odd carbon fatty acids. The even carbon acids give no glucose in phlorizinized dogs or liver glycogen in normals. They are ketogenic. In contrast the odd carbon acids are not ketogenic and do give rise to glycogen in the liver and glucose in phlorizinized dogs. In 1912 Ringer (45) made a quantitative recovery of glucose from injected propionic acid, and it is now believed that

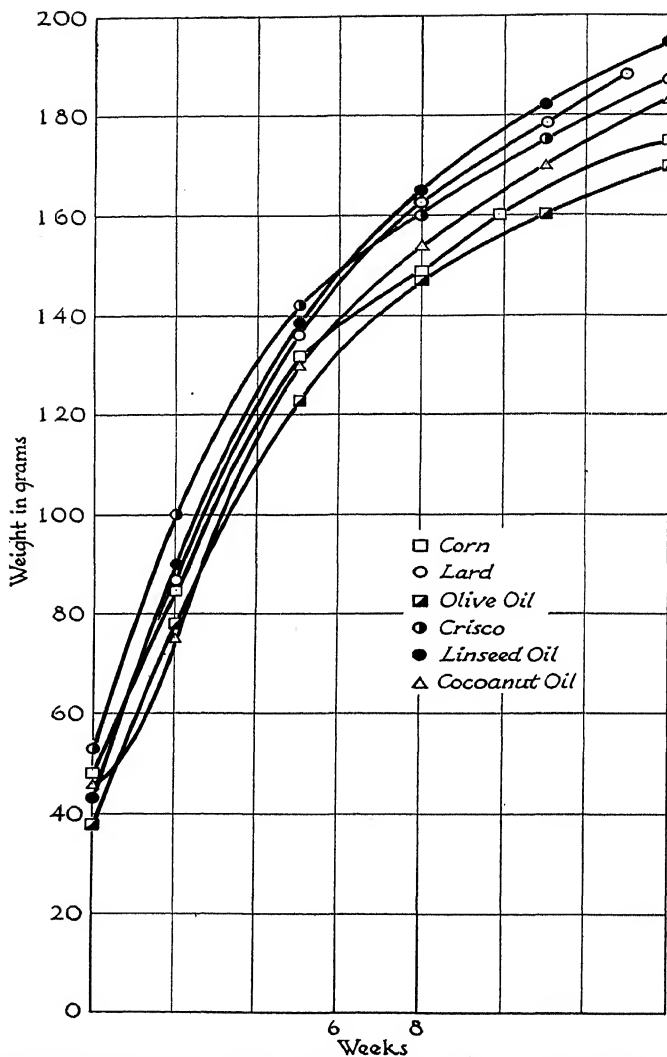


Figure 3. Growth of rats on fats at the 20 per cent level. (Unpublished data of W. R. Brown and G. O. Burr. Drawn in the Medical Art Shop, University of Minnesota.)



## FATS IN THE DIET

the 3 carbon fragment from odd carbon fatty acids prevents ketosis (46), and causes glycogen deposition.

In general the odd carbon acids burn almost as well as the natural even carbons and are not much more toxic at high levels. When used as 10 per cent of the diet, the 7 and 8 carbon acid glycerides (heptylin and caprylin) promote equal growth in rats (see Figure 2). On the other hand, stearin and margarin (17 and 18 carbon) are equally poorly utilized.

*Fatty acid utilization as affected by unsaturation.*—Excepting butter, certain palm oils, and fish oils, food fats from plants and animals are almost wholly composed of fatty acids with 16 and 18 carbons. Most nutrition studies, therefore, are concerned with palmitic, stearic, oleic, linoleic, and linolenic acids. As already indicated, fats that are too hard (saturated) are poorly digestible, and that naturally affects their nutritive value. Aside from those cases, however, there are dozens of edible fats that are more than 97 per cent digestible, varying in iodine number from 60 to 180. It is possible that their food value might be affected by the different proportions of the five above named saturated and unsaturated acids as well as by their arrangement in the glyceride. As a matter of fact the common oils do differ in their effects on rat growth and general health. Of six oils compared at the 20 per cent level in this laboratory the two poorest were olive oil and corn oil, and the two best were lard and linseed oil (see Figure 3). A similar effect has just been shown by Gullickson and Fontaine (47). When fats are homogenized into skim milk and the filled milk fed in place of whole milk, the liquid fats (corn, cottonseed, and soybean oils) have a marked depressive effect on growth, while lard and butter fat gave growth comparable to whole milk. The calves in the vegetable-oil group were emaciated and had rough hair coats. Death frequently occurred. From these two experiments on very different species it would seem that it is possible to feed too much oleic and linoleic acids for the good of the animal.

The cause of such an effect is not clear. It is possible that there is a normal composition of body lipids which must not

## GEORGE O. BURR

be greatly altered if the tissues are to function best. All old work is agreed that ingested fat markedly affects body fat. In 1882 and 1883 Lebedev starved dogs to deplete body fat and then fed linseed oil and mutton fat. The former produced a reserve fat which melted below  $0^{\circ}$ ; the latter, one which melted above  $50^{\circ}$ . The recent work of Barbour (48) indicates that it is very difficult or impossible to deposit much more than the nor-

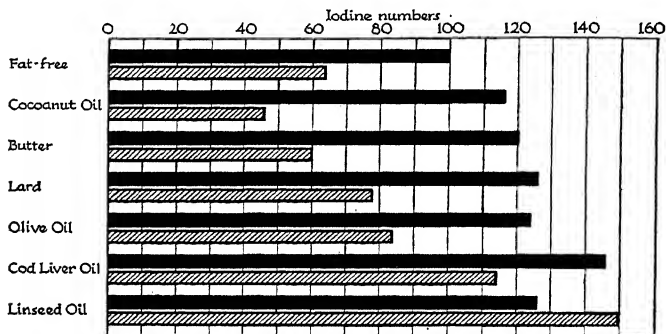


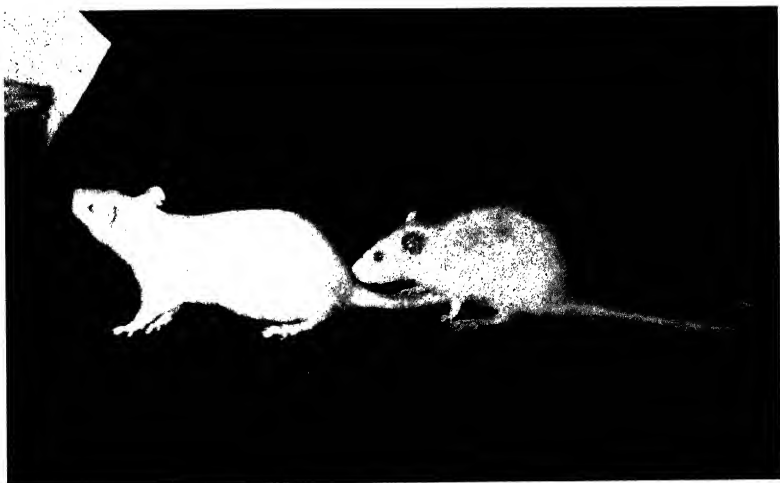
Figure 4. The comparative influence of various fats on the degree of unsaturation of the phospholipid and neutral-fat fatty acids from the entire bodies of rats. (From Sinclair. Drawn in the Medical Art Shop, University of Minnesota.) Solid bars: phospholipid fatty acids. Hatched bars: neutral-fat fatty acids.

mal 25 per cent of saturated acids in the body of the rat. There is a threshold, he says, limited by excretion and burning. On the other hand, body fats may be made extremely soft and abnormal by the ready uptake of unsaturated acids. These effects can be best illustrated by the work of Sinclair (49). Figure 4 shows that the effects of feeding a given fat cannot be predicted, because of a selective action among different tissues and by the phospholipids and neutral fats. The solid bars represent phospholipids, the hatched bars neutral fats. The contrast between cod-liver oil and linseed oil is very marked. The phospholipids readily take up the highly unsaturated acids from the former and reject the linolenic acid from the latter.

It is suggested by McCay, Paul, and Maynard (50) that the



*Figure 5.* Scaly skin of hind feet and tail of rat on fat-deficient diet.



*Figure 6.* Litter mate sisters with and without fat in the diet.



## FATS IN THE DIET

highly unsaturated high molecular weight fatty acids of the liver are responsible for the peculiar effects of cod-liver oil. One-third of a cc. per kilo body weight daily markedly reduces butter fat production by cows and goats. An appreciable amount of these special acids appears in the butter (Hilditch, 51). Cod-liver oil has also become famous for its effects on the muscles of herbivores. Four per cent in the diet quickly kills guinea pigs. Partial hydrogenation to an iodine number of 60 prevents the effect both on muscles and on milk fat production.

*Fat synthesis and the essential fatty acids.*—As indicated earlier in this discussion, the clear proof of fat formation from other foods led to a general acceptance of the view that fat is not an essential part of the diet. However, there is just as clear evidence now that this power of synthesis by higher animals is limited and that some fat must be included in the diet. Careful exclusion of fat from the diet of rats leads to (1) development of scaly skin (see Figure 5); (2) marked retardation of growth (see Figure 6); (3) kidney lesions and hematuria (51a; see Figure 7); and (4) early death (see Figure 8).

It has been shown that these effects are due to a lack of linoleic acid, which is now called essential. The failure of pigs to synthesize this acid is illustrated by the work of Ellis and Hankins (52) (see Figure 9). The pigs were fed a low-fat, high-carbohydrate diet. At killing time their bodies contained much more fat than was consumed in the feed. On comparing the individual fatty acids it was found that synthesis was limited to oleic and saturated acids. In 230-pound pigs there had been an accumulation of 34 pounds of synthetic oleic and 28 pounds of saturated acids. Synthesis must have been even greater, since between meals some synthetic fat is burned.

In marked contrast is the figure for linoleic acid, which is so constantly consumed in metabolism that there are two and one-half pounds less in the body than is eaten in the diet. For good growth of the rat probably 20 mg. of linoleic acid is required daily. There is no evidence of the production of linoleic acid by desaturation nor is there proof that no linoleic acid can



Figure 7. Photomicrograph of a section of the kidney from test rat 3, showing necrosis, apical disintegration, and calcification of the papilla; also hyperplasia of the pelvic epithelium. Hematoxylin-eosin stain; x 20; d, papillary duct; c, calcification; epi, proliferated pelvic epithelium; f, perirenal fat. (From Borland and Jackson.)

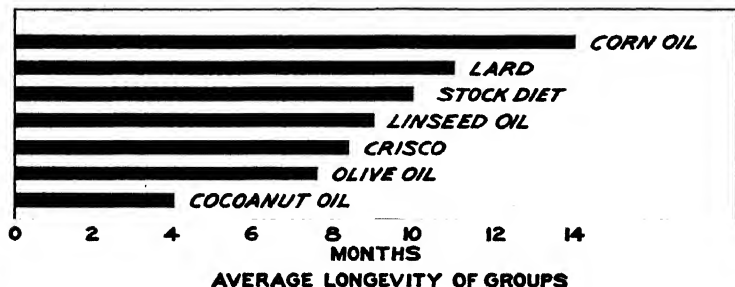


Figure 8. The average longevity of rats reared on a fat-deficient diet depends on the diet of the mother. The mothers were reared on simplified diets containing 20 per cent of one of the above fats. (Unpublished data of W. R. Brown and G. O. Burr.)

## FATS IN THE DIET

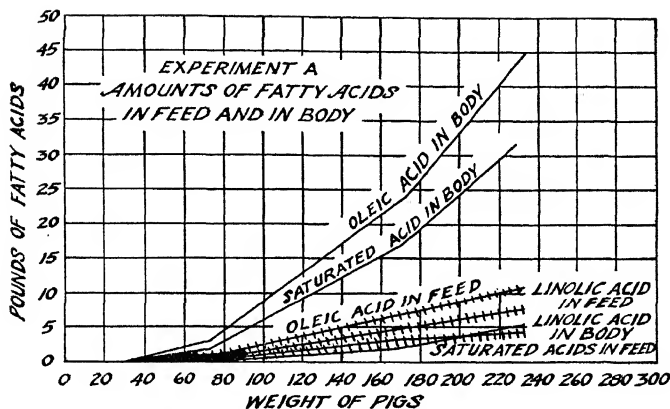


Figure 9. Total fatty acids synthesized by pigs on a low-fat diet. (From Ellis and Hankins.)

be produced by higher animals. Linoleic acid is considered by us as comparable to the essential amino acids. All muscle and organ tissue contains a large amount of this acid. It is an essential part of the cell and cannot be reduced by starvation below a certain minimum. The iodine number of rat phospholipids cannot be reduced below 100, which indicates an appreciable amount of linoleic acid (22). In lecithin from normal beef liver every third fatty acid is linoleic (53).

Some years ago (54) we called attention to the fact that cod-liver oil caused renewed growth of fat-deficient rats without completely freeing the skin from scale. Later a comparison of corn oil, linseed oil, and cod-liver oil made it clear that whereas linseed oil gives superior growth it does not cure the skin as well as corn oil. Cod-liver oil shows an even greater differential. At the 20 mg.-daily level the growth effect is equal to that of corn oil, but there is no relief of the skin trouble (see Figure 10).

A widely accepted but erroneous analysis of linseed oil led us to believe that the differences between corn oil and linseed oil were due to the presence of isomeric forms of linoleic acid, the

## GEORGE O. BURR

so-called alpha and beta. As a result of work here and in other laboratories we are convinced that there is only one form of linoleic acid in most plant fats and that the differences between corn oil and linseed oil are due to the different curative effects of linoleic and linolenic acids. Very recent analyses of corn oil and linseed oil show that corn oil is about 60 per cent linoleic acid while linseed oil has about 40 per cent each of linoleic and linolenic. The results in Figure 10 would be explained if linoleic and linolenic acids had about equal growth effects while only linoleic cured the skin.

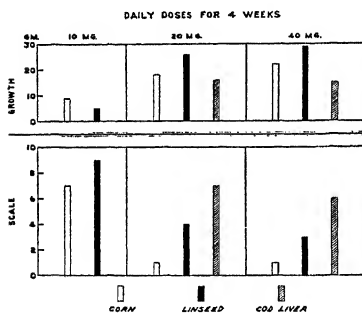


Figure 10. Growth and skin response to corn, linseed, and cod-liver oils.

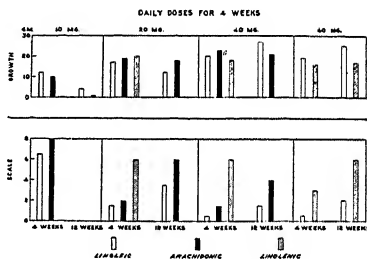


Figure 11. Growth and skin response to linoleic, arachidonic, and linolenic acids. Scaliness estimated visually; the higher the column the poorer the skin. (Unpublished data, J. P. Kass, J. B. Brown, G. O. Burr.)

Figure 11 shows that this is actually the case. Linolenic acid gives an excellent growth response but almost no skin cure, even at the 40 mg. level. Arachidonic acid, on the other hand, aids both growth and skin to about the same extent as linoleic. The chief difference noticed is that when a large dose (40 mg.) is fed, the storage effect of the linoleic becomes evident, so that growth continues and the skin stays clear longer after the doses are discontinued. Studies are being continued on the mode by which arachidonic and linolenic can substitute wholly or partially for linoleic. Of all the fatty acids fed, including numerous isomers and so-called isomers, only arachidonic approaches linoleic in effectiveness in the cure of both growth and scale. It is



## FATS IN THE DIET

evident from Table 2 that space configuration and double bond position is important. It is possible that in the elaidinized linolenic there is one isomer which still has two of the double bonds in the right space configuration to be effective. The fair growth produced by Hilditch's concentrate of butter iso-linoleic may be due to contamination with ordinary linoleic. Further study on this fraction is required.

TABLE 2. — SPECIAL PREPARATIONS AND ISOMERS WHICH HAVE BEEN TESTED  
(UNPUBLISHED DATA OF J. P. KASS, J. B. BROWN, AND G. O. BURR.)

Group	Substance Fed	Response
Pure compounds:		
125.....	Linolelaidic acid	Negative
116.....	9, 11-linoleic acid	Negative
115.....	Eleostearic acid	Negative
Mixtures containing known compounds:		
121.....	Selenium elaidinized linoleic	Very poor
121.....	Nitrate elaidinized linoleic	Negative
125.....	Selenium elaidinized linolenic	Fair growth
115-x } 118-A } 125 }	Alkali treated olive oil, corn oil, linseed oil, linoleic, and linolenic (contains 10, 12-linoleic, 10, 12, 14-linolenic, and other isomers, but no ordinary linoleic or linolenic)	Negative
121.....	Castolene (dehydrated castor oil, contains 9, 11-linoleic and possibly other isomers)	Very poor
115.....	Debrominated liquid bromides (J. B. B.*)	Fair growth
118-x.....	Debrominated liquid bromides (J. P. K.†)	Fair growth
121.....	Butter concentrate (30% iso-linoleic, T. P. H.‡)	Fair growth

\* We are indebted to J. B. Brown for this sample.

† Prepared by J. P. Kass.

‡ We are indebted to T. P. Hilditch for this sample.

*Practical application.*—Numerous practical results have come from the consideration of dietary fat. It has been shown by Maynard, Gardner, and Hodson (55) that when the diet of cows contains less than 4 per cent fat there is a decrease in both milk and butter fat production. It may be brought back to normal by the inclusion of common fats like soybean oil. This has become an important problem in recent years with the addition of molasses, beet pulp, and other low-fat foods to stock feeds.

## GEORGE O. BURR

One of the most striking results of the observation on scaly skin production is the demonstration that some of the common skin troubles can be helped by feeding moderate doses of oils like linseed, corn, or lard.

It was found that the serum fatty acids of fat-deficient rats had low iodine numbers (56) as compared with normals on stock diet. The average values were 107 and 136 respectively.

The same relationship was found by Hansen for eczematous infants as compared with normals (57) (average iodine numbers 84 and 111 respectively). Further studies indicate that both linoleic and arachidonic acids are low in the serum of these infants (58). When these infants are fed daily doses of unsaturated fats there is a rise in the iodine number of the serum lipids and a simultaneous improvement in the clinical condition (59). Similar good effects have been reported from several laboratories.

It now appears that vitamin B deficiency affects the animal specifically through its failure to metabolize pyruvic acid. This substance is an intermediate in the metabolism of carbohydrates but not of fat. This fact gives an explanation of the early observation of Funk (60) that the severity of polyneuritis could be correlated with the amount of carbohydrate in the diet. Quantitative studies of this problem by Evans and coworkers and others have shown that with diets of fat and protein alone no thiamine is required. Even 25 per cent of fat in the diet reduces the vitamin requirement tremendously. It is likely that with moderately high fat diets there would be no human beriberi.

The above effect is general for all fats, although there are small differences among them. With the more recently discovered vitamin B<sub>6</sub> there is also a fat-sparing action, but in this case the effect is due to unsaturated acids. The outward symptom of B<sub>6</sub> deficiency is a skin lesion called acrodynia. It has been found that acrodynia cannot be produced on high-fat diets, and a published list of oils giving their effectiveness in preventing this syndrome shows that the value of the oil is proportional to its linoleic acid content (61). Hence B<sub>6</sub> cannot

## FATS IN THE DIET

prevent fat deficiency, but fat can prevent B<sub>6</sub> deficiency as far as the acrodynia is concerned.

In conclusion I should like to point out that there are fatty acids in tissues which have not yet been identified, and we do not even have adequate analytical methods for the known ones.

The problem of fatty livers and the effects of choline, filtrate factor, and other substances on fat mobilization is being intensively studied by Best, McHenry, and others, but much is yet to be learned. In 1936 Smedley-MacLean (62) stated that hardly anything is known or explained about the process of fat synthesis; and as Barron (63) pointed out this year, our knowledge of the enzymes producing the oxidation of fats is nil. In view of these facts it seems safe to say that the science of fatty acid metabolism is in its infancy.

### *References*

1. Sinclair, R. G. 1934. *Physiol. Rev.* 14:351.
2. Bills, C. W. 1935. *Physiol. Rev.* 15:1.
3. Terroine, E. F. 1936. *Ann. Rev. Biochem.* 5:227.
4. Anderson, W. E., and Williams, H. H. 1937. *Physiol. Rev.* 17:335.
5. Verzár, F. 1938. *Ann. Rev. Biochem.* 7:163.
6. Leathes, J. B. 1908-09. *Harvey Lectures* 4:213.
7. Terroine, E. F. 1919. *Ann. d. sci. nat. Zool.* 4:s.40,1.
8. Stepp, Wilhelm. 1919. *Zeitschr. f. Biol.* 57:135.
9. Osborne, T. B., and Mendel, L. B. 1912. *Jour. Biol. Chem.* 12:81.
10. Aron, H. 1918. *Biochem. Zeitschr.* 92:211.
11. Salkowski, E. 1919. *Biochem. Zeitschr.* 94:205.
12. Von Gröer, F. 1919. *Biochem. Zeitschr.* 97:311.
13. McCollum, E. V., and Davis, M. 1914. *Jour. Biol. Chem.* 19:245.
14. Drummond, J. C., and Coward, K. H. 1921. *Lancet* 201:698.
15. Kohman, E. A. 1920. *Am. Jour. Physiol.* 51:378.
16. Osborne, T. B., and Mendel, L. B. 1920. *Jour. Biol. Chem.* 45:145.
17. Krogh, A., and Lindhard, J. 1920. *Biochem. Jour.* 14:290.
18. Hindhede, M. 1920. *Skand. Arch. f. Physiol.* 39:78.
19. Murlin, J. R. 1918. *Jour. Am. Med. Assn.* 71:950.
20. Levine, H., and Smith, A. H. 1927. *Jour. Biol. Chem.* 72:223.
21. Bloor, W. R. 1924. *Harvey Lectures* 19:39.
22. Sinclair, R. G. 1929. *Jour. Biol. Chem.* 82:117.
23. Sullman, H., and Willbrandt, W. 1934. *Biochem. Zeitschr.* 270:52.
24. Verzár, F., and Laszt, L. 1934. *Biochem. Zeitschr.* 270:24, 35.
25. Miller, E. S., Barnes, R. H., Kass, J. P., and Burr, G. O. In press. *Proc. Soc. Exper. Biol. Med.*

# GEORGE O. BURR

26. Bloor, W. R. 1939. *Physiol. Rev.* 19:557.
27. Langworthy, C. F. 1923. *Jour. Indust. Engin. Chem.* 15:276.
28. Tofte, F. 1934. *Biochem. Zeitschr.* 272:308.
29. Steenbock, H., Irwin, M. H., and Weber, J. 1936. *Jour. Nutr.* 12:103.
30. Verzá, F. 1932-33. *Nutr. Abstr. Rev.* 2:441.
31. Holt, L. E., Jr., Tidwell, H. C., Kirk, C. M., Gross, D. M., and Neale, A. S. 1935. *Jour. Pediat.* 6:427.
32. Verzá, F., and McDougall, E. J. 1936. *Absorption from the Intestine.* Longmans, Green and Co., New York.
33. Ozaki, J. 1927. *Biochem. Zeitschr.* 189:233.
34. Evans, H. M., and Lepkovsky, S. 1933. *Jour. Biol. Chem.* 96:165.
35. Barbour, A. D. 1933. *Jour. Biol. Chem.* 101:63.
36. Lepkovsky, S., Ouer, R. A., and Evans, H. M. 1935. *Jour. Biol. Chem.* 108:431.
37. Cox, W. M. 1933. *Jour. Biol. Chem.* 103:777.
38. Davis, R. E. 1930. *Jour. Biol. Chem.* 88:67.
39. Eckstein, H. C. 1929. *Jour. Biol. Chem.* 81:613.
40. ——— 1929. *Jour. Biol. Chem.* 84:353.
41. Powell, M. 1930. *Jour. Biol. Chem.* 89:547.
42. ——— 1932. *Jour. Biol. Chem.* 95:43.
43. Quick, A. J. 1928. *Jour. Biol. Chem.* 77:581.
44. Verkade, P. E. 1938. *Jour. Soc. Chem. Indust.* 57:704.
45. Ringer, A. I. 1913. *Jour. Biol. Chem.* 14:43.
46. Deuel, H. J., Jr., Butts, J. S., Hallman, L. F., and Cutter, C. H. 1935. *Jour. Biol. Chem.* 112:15.
47. Gullickson, T. W., and Fontaine, F. C. 1939. *Jour. Dairy Sci.* 22:471.
48. Barbour, A. D. 1934. *Jour. Biol. Chem.* 106:281.
49. Sinclair, R. G. 1931. *Jour. Biol. Chem.* 92:245.
50. McCay, C. M., Paul, H., and Maynard, L. A. 1938. *Jour. Nutr.* 15:367.
51. Hilditch, T. P., and Thompson, H. M. 1936. *Biochem. Jour.* 30:677.
- 51a. Borland, V. G., and Jackson, C. M. 1931. *Arch. Path.* 11:687.
52. Ellis, N. R., and Hankins, O. G. 1925. *Jour. Biol. Chem.* 66:101.
53. Snyder, R. H., and Bloor, W. R. 1933. *Jour. Biol. Chem.* 99:555.
54. Burr, G. O., Burr, M. M., and Brown, W. R. 1931. *Proc. Soc. Exper. Biol. Med.* 28:905.
55. Maynard, L. A., Gardner, K. E., and Hodson, A. 1939. *Cornell Univ. Agr. Exper. Sta. Bull.* Vol. 722.
56. Hansen, A. E., and Burr, G. O. 1933. *Proc. Soc. Exper. Biol. Med.* 30:1201.
57. Hansen, A. E. 1933. *Proc. Soc. Exper. Biol. Med.* 30:1198.
58. Brown, W. R., and Hansen, A. E. 1937. *Proc. Soc. Exper. Biol. Med.* 36:113.
59. Hansen, A. E. 1937. *Am. Jour. Dis. Child.* 53:933.
60. Funk, C. 1914. *Zeitschr. f. physiol. Chem.* 89:378.
61. Schneider, H. A., Ascham, J. K., Platz, B. R., and Steenbock, H. 1939. *Jour. Nutr.* 18:99.
62. Smedley-MacLean, Ida. 1936. *Ergebn. Enzymforsch.* 5:285.
63. Barron, E. S. G. 1939. *Physiol. Rev.* 19:184.

# HEPARIN AND THROMBOSIS

BY

Charles H. Best

*Professor of Physiology,  
University of Toronto*

I KNOW that you will appreciate fully that our researches, the results of which have little immediate application under our present circumstances, have lost much of their former charm. The subject for today's lecture is more satisfactory from my own point of view than others I might well have chosen when I accepted your committee's invitation several months ago. Heparin promises to be of considerable value in numerous phases of the investigations now commanding our attention.

We shall certainly never know the name of the individual who first watched the obvious changes taking place in clotting blood. It is difficult to be sure who made the first scientific observations. Aristotle noted that the blood did not coagulate when the fibers were removed and therefore probably ascribed the process to the presence of the fibrous material. We find notes on this phenomenon of clotting in the writings of a great many of the famous contributors to medicine, among them Harvey, Willis, Malpighi, Lower, Borelli, Boyle, Haller, Sydenham, and Leeuwenhoek. The last mentioned watched the clotting process under the microscope but did not recognize any of the essential details of the process. Richard Davies and William Hewson, in the middle seventeen hundreds, recognized the essential role of fibrin in clot formation and discussed the change from "liquid" to solid fibrin.

This very brief and incomplete reference to blood clotting is merely intended as an introduction to the subject of thrombosis. A blood clot is, of course, a thrombus. When the thrombus is made up entirely of clot it is appropriately termed a *red thrombus*. On the other hand a thrombus, as will be illustrated later,

may be almost pure white in color and made up completely of platelets. Thus we may have a white thrombus, a red thrombus, or one made up of both platelets and clot, i. e., a mixed thrombus.

The third cell, the platelet or thrombocyte, was probably first seen by Leeuwenhoek. The first systematic study was conducted by Zimmermann (1), who used a solution of magnesium sulfate to prevent clotting of the blood. Virchow, the famous German pathologist, denied that these small cells were constantly present in the blood. Sir William Osler's first publication on the platelets indicated that he thought these cells were the product of bacterial action; but he soon changed his mind, and in 1873 recognized them as definite elements of the blood. He showed that the so-called Schultze's granular masses were made up of many individual platelets. Osler (2) was probably the first to see the platelets within the blood vessel but Bizzozero (3) in 1882 studied this field much more thoroughly and was undoubtedly the first to see the platelets building up into a thrombus. Bizzozero was the first to study the rate of formation of platelets. He removed blood from animals and after separating the platelets reinjected the blood. Bizzozero's observations on thrombus formation marked the beginning of the histological work of Eberth and Schimmelbusch (4), Aschoff (5), Welch (6), Zurbelle (7), and others on the mechanism of thrombus formation, the results of which are given in many textbooks of pathology (see Figures 1 and 2, page 126). It is, of course, well established that damage to the lining of the blood vessels and slowing of the blood flow are two of the important predisposing causes of platelet agglutination. There are certainly other factors.

I shall not discuss the origin of the blood platelets in the body but I should like to refer you to an excellent article by Tocantins (8) for a comprehensive review of this subject. With regard to the role of platelets in blood clotting, while there is no doubt that they contain a kinase, the presence of a heparin-like substance has also been reported (9). It would appear more

## HEPARIN AND THROMBOSIS

likely to me that the platelets become attached to the strands of the fibrin as soon as they are formed, rather than that rupture of the platelets initiates the formation of the strands of fibrin; but there is a good opportunity for further research on this point.

*Heparin.*—While there is an earlier report by Schmidt (10) that a potent anticoagulant could be obtained in certain tissues, McLean (11), in Howell's laboratory, was the first to obtain a useful preparation from normal tissues. Howell and Holt (12) explored the physiological action and the chemical properties of this material and a very great deal of subsequent work in this field was anticipated by Professor Howell. He pioneered in the purification of heparin, in the study of its mode of action, in its use in transfusions of blood between human subjects, and in many other aspects of the problems made obvious by his own discoveries. A review of the literature on heparin has recently been compiled by Mason (13).

In 1929 it appeared to me (and undoubtedly to many others interested in this field) that two things in particular were lacking, which prevented the further exploration of the physiological role of heparin and its clinical application. The first was a product of satisfactory purity and potency; the second, proof that heparin prevented thrombosis as well as coagulation. The methods then available did not, in the hands of my chemical colleagues, provide us with satisfactory material. The yields were extremely small. The best product commercially available was very crude. I was able to interest a young organic chemist, Dr. Arthur Charles, in the problem of purification; and, to make a long story very short, he and Dr. D. A. Scott (14) finally evolved a method which gave us adequate amounts of highly purified heparin. The source was beef lung. While dog's liver contains more heparin than dog's lung, beef lung was found to be a richer source than beef liver. Charles and Scott (15) eventually prepared heparin as a crystalline barium salt, and others have confirmed this finding. This material, after removal of the barium, has been used in all recent clinical and experimental

## CHARLES H. BEST

work in Toronto and in much of that conducted in the United States.

Before summarizing our present knowledge of the chemistry of heparin, I should like to pay tribute to the brilliant work of Jorpes and his colleagues in Stockholm. Jorpes undoubtedly realized, as we did, that the heparin situation would be greatly improved if adequate amounts of a purified product could be made available. Utilizing the basic procedure elaborated by Charles and Scott, he developed a method for the production of purified heparin. The use of the purified material in clinical problems was begun at about the same time in Stockholm and Toronto, and an opportunity of comparing notes has recently been afforded by Dr. Clarence Crafoord's visit to Toronto and Dr. D. W. G. Murray's to Stockholm.

*The chemistry of heparin.*—I am indebted to my colleague Dr. Charles for help in preparing the following brief account of the chemistry of heparin. After Howell and Holt prepared heparin in a crude form, Howell (16) succeeded in obtaining a very active preparation which he believed to be a glycuronic acid derivative. Charles and Scott (15), although confirming the carbohydrate nature of heparin, were unable to obtain a positive test for glycuronic acid even when using very active heparin preparations. This highly active heparin was shown for the first time to contain about 2 per cent of nitrogen. Jorpes (17) made a notable contribution when he showed that the anticoagulant was a sulfuric ester of a carbohydrate. He suggested that the carbohydrate constituents were a hexosamine and a uronic acid (18). Charles and Scott in 1936 reported the isolation of a crystalline barium salt of heparin. They confirmed the finding of Jorpes that heparin is a polysulfuric ester of a carbohydrate, and found that this crystalline substance also contained nitrogen. Later Jorpes (19) succeeded in isolating glucosamine from hydrolyzed heparin. The question of the identity of the uronic acid is still open, but the amount of carbon dioxide evolved when heparin is treated with strong hydrochloric acid does indicate the presence of some uronic acid.



## HEPARIN AND THROMBOSIS

Failing the isolation of the uronic acid in a crystalline form, the naphthoresorcinol test has been used to identify glycuronic acid. Charles and Scott (15) have failed to obtain a positive test with naphthoresorcinol and heparin. Jorpes has on occasions been able to obtain indications of the presence of glycuronic acid in heparin, but not at all consistently, as may be seen from reference to his publications.

Dr. Charles has obtained the crystalline barium salt of heparin from beef lung, beef liver, and beef intestine. Although the purification was carried out in different ways, the crystals were identical in appearance, chemical analysis, and physiological activity. Analytical figures show that the ratio of S:N in the crystals is 5:2. Also an ammonium salt can be prepared from these crystals with an  $\text{NH}_4$ :S ratio of 1:1 and an S:bound N ratio of 5:2. This ammonium salt can be quantitatively reconverted into the barium salt identical in composition with the original crystals. These findings are further confirmation that the crystals are a chemical entity.

Heparin may, in brief, be considered as a carbohydrate complex of unknown molecular size, containing probably a uronic acid and glucosamine—the latter partially, or totally, acetylated with respect to the N-containing group. The hydroxyl groups of the complex are in part or totally esterified with sulfuric acid in such a way that the N:S ratio is 2:5.

*The site of formation of heparin.*—As we have seen, heparin was first prepared from liver, but Howell had also obtained small amounts from muscle, lymph nodes, and blood. In 1933 Charles and Scott (20), using their method, which gave better yields, isolated highly active heparin from lungs, spleen, and kidneys. The lung was found to be particularly rich. More recently Jaques and Charles have secured relatively large yields from intestinal mucosa. Thus lung, liver, and intestine appear to be particularly rich in the anticoagulant.

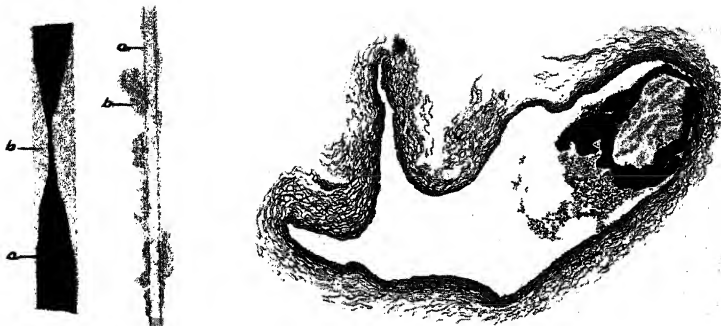
The histological and chemical contributions of Jorpes, Holmgren, and Wilander (21) to our knowledge of the source of heparin in the body and the physiological function of the mast

## CHARLES H. BEST

cells of Ehrlich promise to make a very bright page in the history of the heparin researches. The function of these cells has been shrouded in mystery ever since their discovery by Paul Ehrlich in 1877. The cells are characterized by the metachromatic staining of their granules by basic aniline dyes. These cells have a wide distribution, but many investigators have stressed the fact that they are met with abundantly in connective tissue which is rich in blood vessels. It has been suggested that the mast cells are single-celled glandular organs of the connective tissue (22).

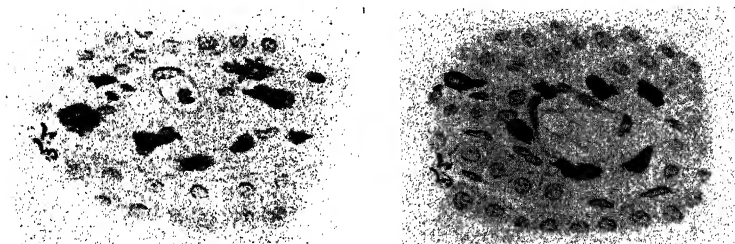
The granules of the mast cells are soluble in water (Ehrlich). Lison (23) has shown that high molecular sulfuric acid esters of polysaccharides produce the metachromatic staining (chondroitin sulfuric acid found in cartilage and elsewhere gives this color). Jorpes showed that heparin gave a similar reaction with toluidine blue, and Bergström (24) noted the same color when the dye was mixed with synthetic sulfuric acid esters of carbohydrate. When toluidine blue was applied by Jorpes to various blood vessels, the characteristic staining was observed, but proof that this was due to heparin was required. Holmgren and Wilander (25), using this dye as a histological stain, found that the color appeared in the mast cells of Ehrlich, which are found mainly in connective tissue near capillaries and in the walls of the larger blood vessels. The granules in the mast cells showed greater affinity for the dye than cells containing chondroitin sulfuric acid, and the tint appeared to be somewhat different. Wilander (26), using the method of Charles and Scott, found a very good correlation between the heparin content of various tissues and the number of mast cells. Thus very strong evidence is available that these cells are the source of heparin in the body. This conclusion is supported by results from another series of investigations which will now be briefly reviewed.

*Liberation of heparin in anaphylactic and peptone shock.*—In 1909 Biedl and Kraus (27) found that the blood became incoagulable in anaphylactic shock. Since that time a great many investigators have studied the problem and almost every factor



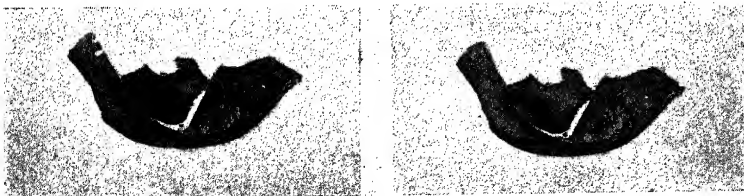
*Figure 1 (left).* Parietal blood platelet thrombus in a stagnating vessel. a: red hyalin thrombus. b: blood platelet thrombus. Hartnack VII. Oc. 3. a: glass filament. b: blood platelets fixed in 1 per cent osmic acid. Hartnack VIII. Oc. 3. (From Eberth and Schimmelbusch.)

*Figure 2 (right).* Thrombus of the jugular vein, 1 hr. old. Thread infected with streptococcus broth. Low magnification. (From Zurhelle.)



*Figure 3 (left).* Mast cells in the liver of a dog after peptone shock. The cells contain sparse granules that are small and for the most part faintly stained. Fixation in 4 per cent basic lead acetate. Stained in 1 per cent aqueous solution of toluidine blue. (From Wilander.)

*Figure 4 (right).* Mast cells in the liver of a normal dog. The cells contain well-fixed granules. The mast cells are grouped around a small blood vessel. Fixation, staining, and magnification as in Figure 3. (From Wilander.)



*Figure 5.* Sections of the tip of the left ventricle 24 hrs. after injury, showing a large adherent mural thrombus. No heparin was given. (From Solandt, Nassim, and Best.)



## HEPARIN AND THROMBOSIS

in the clotting process has been incriminated. Quite recently Eagle, Johnston, and Ravdin (28) have obtained evidence that the antithrombin content of blood was greatly increased. They suspected heparin but did not prove its presence. Waters, Markowitz, and Jaques (29) have shown that the clotting time of blood in anaphylactic shock can be reduced to normal value by the addition of protamine. Protamine neutralizes the effect of heparin by forming with it an inactive salt (30, 31, 32). Waters and Jaques have been able to isolate heparin in crystalline form from the blood of the shocked animals. The amount of heparin in the liver showed a marked fall. The amount appearing in the blood is considerable—an enormous dose would be required to produce the same effect.

The blood in peptone shock may also be incoagulable and here again many theories have been entertained. Howell (33) and Quick (34) have, however, obtained strong evidence in favor of the appearance of heparin during the shock. More recently Waters and Jaques have neutralized with protamine the anticoagulant which appears in the blood, and Wilander (26) has prepared highly active heparin from the blood of animals in peptone shock. Wilander has secured evidence that this anticoagulant comes, in part at least, from the mast cells in the liver (see Figures 3 and 4).

*Administration of heparin.*—If a single dose of heparin be given intravenously, the clotting time after a definite latent period becomes greatly prolonged (see Figure 6). The increase in the clotting time depends on the dose of heparin given. There is no negative phase, i. e., the clotting time does not go below normal after a massive dose of heparin; it comes back fairly accurately to the initial value. Heparin can be given subcutaneously as well as intravenously. The crude material first used was not well absorbed, but purified heparin is absorbed quite rapidly from the tissue spaces and a good effect can be obtained as a result of subcutaneous administration. In the rat the effect of one subcutaneous dose may be to send up the clotting time to over 24 hours, the return to normal occupying from 12 to 24

# CHARLES H. BEST

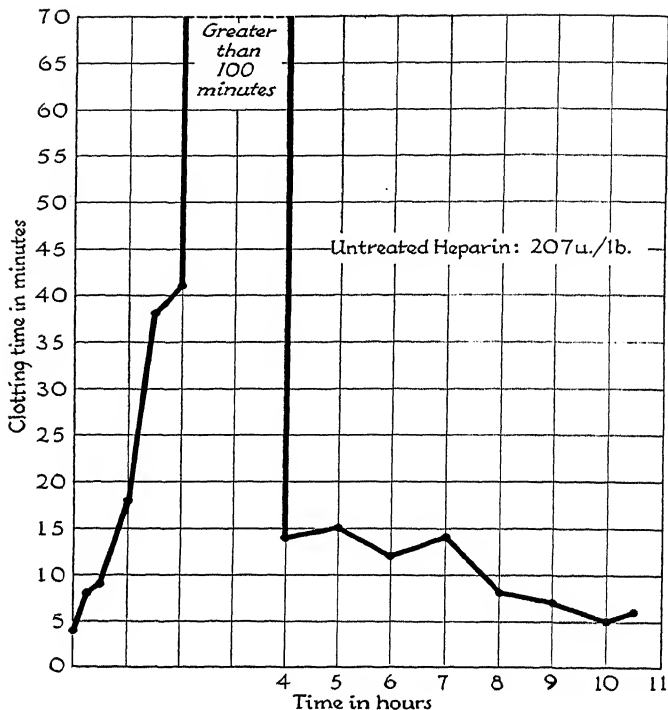


Figure 6. Intravenous injection of heparin. (From Jaques, Charles, and Best. Drawn in the Medical Art Shop, University of Minnesota.)

hours or longer (see Figure 7). However, two out of sixty experimental animals died as a result of hemorrhage into their subcutaneous tissues after the administration of heparin. There was not a great deal of hemorrhage, but 2 or 3 cc. is a large amount of blood for a rat. It is possible that the damage produced locally would not be sufficient to cause a dangerous or even a serious amount of bleeding in large animals.

When heparin is precipitated with protamine a very insoluble compound is formed, which cannot readily be suspended (35).

## HEPARIN AND THROMBOSIS

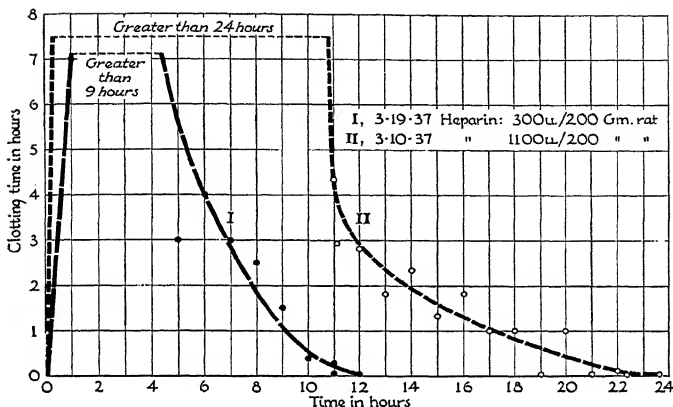


Figure 7. Subcutaneous injection of heparin. (From Jaques, Charles, and Best. Drawn in the Medical Art Shop, University of Minnesota.)

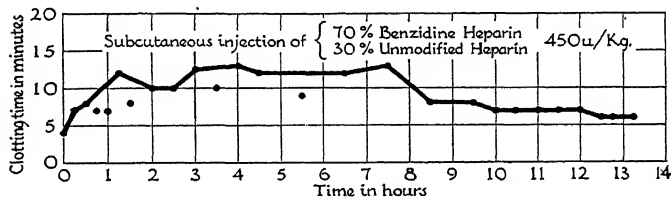


Figure 8. Subcutaneous administration of benzidine-heparin. (From Jaques, Charles, and Best. Drawn in the Medical Art Shop, University of Minnesota.)

With benzidine, however, a compound can be prepared which may be absorbed quite slowly and gives a clotting time of 10–15 minutes for fairly long periods (31) (see Figure 8). It was found advisable to have a little unmodified heparin in the mixture, since certain doses of benzidine-heparin given alone will have no effect at all; but if given with the unmodified heparin the immediate and prolonged action are both forthcoming. It is not suggested, however, that heparin in a combined form should be used in experimental or clinical work. An initial intravenous injection of undiluted heparin (10 mg. per cc.) followed by the

# CHARLES H. BEST

continuous intravenous administration of material diluted in saline provides the best procedure for experimental work.

*Heparin and thrombosis.*—There is a tendency to regard the clotting of blood and the agglutination of platelets as processes governed by the same factors. While there may be much in common between these two they can apparently be completely separated. One has only to watch the clotting of blood from which the platelets have been removed and the formation of a platelet thrombus which contains no detectable amounts of fibrin to realize the independence of these phenomena. It therefore appeared that adequate proof would have to be provided that heparin prevented thrombosis, i. e., platelet agglutination, before it could be recommended for clinical trial. The results of experiments conducted by Shionoya (36) suggested that heparin was not effective in preventing agglutination of platelets; but, as Professor Howell suggested at the time, there was a strong possibility that more highly purified heparin would exert an effect. The researches of our group may be divided into four parts. In the first investigation Dr. Gordon Murray of the Department of Surgery and I, in collaboration with Mr. Jaques and Dr. Perrett (37), found that thrombi forming on the intimal surface of veins which had been injured by mechanical or chemical means could be completely prevented by the adequate administration of purified heparin (see Table 1). It was further

TABLE 1. — EFFECT OF PURIFIED HEPARIN ON INTIMAL SURFACE OF VEINS INJURED BY MECHANICAL OR CHEMICAL MEANS \* (FROM MURRAY, JAKES, PERRETT, AND BEST)

Experiment	Control		Heparin		Heparin u/lb./hr.	Hrs. Heparin Given
	Occluded	Patent	Occluded	Patent		
68 .....	1	1				
69 .....	2	0				
30 .....	4	0				
72 .....	2	0	0	2	10.3	68
73 .....	0	0	0	4	10.9	73
74 .....	1	1	0	2	10.5	72
75 .....	0	0	0	4	10.0	72
Total .....	10 (83%)	2 (17%)	0	12 (100%)		

\* 0.15 cc. soricin injected in all veins.



## HEPARIN AND THROMBOSIS

noted that after three days of continuous heparinization the vein was completely healed and there was no indication that a thrombus would form subsequently at the site which had been injured. In a later paper (38) the formation of white thrombi was studied. It was found that in dogs, cats, and monkeys, the agglutination of platelets could be inhibited or prevented when large amounts of heparin were employed. It may be remarked here that much higher concentrations of heparin are required to prevent agglutination of platelets in glass cells than in the body. In the third series of investigations Dr. D. Y. Solandt and I (39) produced coronary thrombosis by isolating the coronary artery and injecting sodium ricinoleate within the lumen. This material was kept in contact with the intima for five minutes and the clamps on the vessel were then released. In almost every case when no heparin was used a thrombosis was present and in the heparinized series this was almost never observed. The fourth series has just been completed by Dr. Solandt, Dr. Nassim, and myself (40). This had to do with the production of cardiac mural thrombi and the prevention of their formation by the administration of heparin. A technique was evolved by which large mural thrombi could be regularly produced in the lumen of the left ventricle. The endocardium of this cavity was injured by injecting sodium ricinoleate and the myocardium was damaged by ligating the anterior descending branch of the left coronary artery (see Figure 5, page 126). It is clear that without heparin there was a rapid formation of thrombus, but none was seen in those experiments in which the heparin was given well before the injury was produced (see Table 2). The results of all these experiments leave no doubt that under certain experimental conditions the effect of heparin in preventing thrombosis can be readily demonstrated.

(A colored film was then projected, the first part of which showed the flow of blood through a glass cell introduced between the carotid artery and the jugular vein of the monkey. The rapid formation of white thrombi was well illustrated, and the breaking off of large masses to form emboli. The growth of

# CHARLES H. BEST

the thrombus downstream could be clearly seen. The thrombi tended to form in the periphery of the cell, where the flow is slowest, but also on a scratch in the axial stream, where the flow was very rapid. The action of heparin in preventing the formation of white thrombi was demonstrated. In the second part of the film the passage of a plasma containing platelets but very few other cells through a small glass perfusion chamber was shown. The clumpings of the platelets could be clearly seen, together with the formation of the digitations.)

*Clinical applications.*—An adequate supply of highly purified heparin has made possible many types of clinical investigation. In Toronto (41) and in Stockholm (42) patients who have undergone operative procedures, which are followed by a rela-

TABLE 2.—MURAL THROMBUS EXPERIMENTS (24 HOUR)  
(FROM SOLANDT, NASSIM, AND BEST)

(Exp.: Left desc. coronary artery tied; 2 cc. 5 per cent soricin infiltrated around tip.)

Date	Dog	Weight (pounds)	Ekg R-wave	Shortest Clotting Time Observed	Thrombus	
					Macroscopic	Microscopic
CONTROL SERIES — NO HEPARIN						
Nov. 29 ....	12	14	.....		Large *	} Large (white, mixed, and red)
Nov. 29 ....	13	20	.....		Medium*	
Nov. 29 ....	14	..	.....		Medium*	
Nov. 29 ....	15	..	.....		Medium*	
Nov. 30 ....	16	32	.....		Large *	
Nov. 30 ....	17	22	.....		Medium *	
Dec. 1.....	18	23	.....		Medium *	
Dec. 1.....	20	23	.....		Large *	
Dec. 1.....	21	25	.....		Large *	}
Dec. 6.....	22	20	1/32"		Large *	
HEPARIN SERIES †						
Dec. 22.....	51	32	.....	1½ hrs.	} No mural thrombus or clot	
Dec. 22.....	52	31	.....	1½ hrs.		
Dec. 27.....	53	14	.....	over 5 hrs.‡		
Dec. 29.....	55	18	.....	1¾ hrs.		
Jan. 5.....	61	18	Not done	over 3½ hrs.‡		
Jan. 9.....	64	17	Not done	10 min.		

\* Mural thrombus, white, mixed, and red.

† 70 u/kg. given 3–5 min. before soricin injection, and second similar dose approximately ¾ hr. later. Continuous intravenous of heparin at the rate of 40 u/kg./hr. started during this ¾-hr. interval and continued until autopsy.

‡ No early sample.

## HEPARIN AND THROMBOSIS

tively high incidence of embolic phenomena, have been heparinized during the so-called danger period. While it will perhaps be almost impossible to secure scientific proof of the effects of heparin under these conditions, the results have not been unfavorable. In many individual cases it is difficult to refrain from the conclusion that further thrombus formation has been prevented.

*Heparin in vascular surgery.*—The results of Murray's work in Toronto and that of various surgeons in Sweden leave little doubt that heparin is a valuable adjuvant in certain types of vascular surgery, e. g., embolectomy, vascular anastomosis, etc.

*Heparin in transfusions.*—This material has not been used extensively for blood storage. In Sweden and in England heparinization of donors has been studied. The intravenous injection of 1 or 2 cc. of heparin (10 mg. per cc.) will raise the donor's clotting time to 20 or 50 minutes and will thus facilitate bleeding, if difficulties in clotting or thrombus formation have been encountered.

*Heparin in thrombosis of the retinal vessels.*—The only place in the human body where the process of thrombus formation can readily be watched is in the retinal vessels. Promising results with the use of heparin have been reported by Holmin (43), Ploman (44), and by other Swedish workers in the treatment of thrombosis of the central vein of the retina. While these clinical findings appear to demonstrate the favorable effect of heparin, many more cases must be studied before the matter is settled.

*Heparin in subacute bacterial endocarditis.*—It may be considered that this subject should not be mentioned at this stage of the development. The possibility that the prevention of fibrin and thrombus formation might render the *Streptococcus viridans* more vulnerable to attack is obvious. The idea that such a study might be interesting crossed my mind early in the heparin investigations. A little later it was suggested to me by Dr. William Thalhimier. However, nothing was done at the time, and the first attempt to modify the course of this disease

## CHARLES H. BEST

by use of heparin was made by Katz and his colleagues in Chicago (45).

Recently there has been great interest in the use of sulfapyridine and heparin. A brief preliminary report has been published by Kelson and White (46). They emphasize the fact that many more cases must be observed before it can be stated that the use of heparin represents an advance over that of the sulfanilamide derivatives alone.

### References

1. Zimmermann, G. 1860. Zur Blutkörperchenfrage. *Virchows Arch. f. path. Anat.* 18:221-42.
2. Osler, W. 1886. Cartwright lectures on certain problems in the physiology of the blood corpuscles. I. The blood plaque. *Med. News* 48: 365-70.
3. Bizzozero, J. 1882. Ueber einen neuen Formbestandteil des Blutes und dessen Rolle bei der Thrombose und der Blutgerinnung. *Virchows Arch. f. path. Anat.* 90:261-332.
4. Eberth, J. C., and Schimmelbusch, C. 1886. Experimentelle Untersuchungen über Thrombose. *Virchows Arch. f. path. Anat.* 103:39-87; 105: 456-65.
5. Aschoff, L. 1924. *Lectures on Pathology*. Hoeber, New York. pp. 253-78.
6. Welch, W. H. 1920. *Papers and Addresses*. Johns Hopkins Press, Baltimore. 1:110-92.
7. Zurhelle, E. 1910. Experimentelle Untersuchungen über die Beziehungen der Infektion und der Fibringerinnung zur Thrombenbildung im strömenden Blut. *Beitr. z. path. Anat. u. z. allg. Path.* 47:539-70.
8. Tocantins, L. M. 1938. The mammalian blood platelet in health and disease. *Medicine* 17:155-260.
9. Chargaff, E., Bancroft, F. W., and Stanley-Brown, M. 1936. Studies on the chemistry of blood coagulation. III. The chemical constituents of blood platelets and their role in blood clotting, with remarks on the activation of clotting by lipids. *Jour. Biol. Chem.* 116:237-51.
10. Schmidt, A. 1892. *Zur Blutlehre*. Leipzig.
11. McLean, J. 1916. The thromboplastic action of cephalin. *Am. Jour. Physiol.* 41:250-57.
12. Howell, W. H., and Holt, E. 1918. Two new factors in blood coagulation — heparin and pro-antithrombin. *Am. Jour. Physiol.* 47:328-41.
13. Mason, M. F., 1939. Heparin: a review of its history, chemistry, physiology and clinical applications. *Surgery* 5:451-67, 618-37.
14. Charles, A. F., and Scott, D. A. 1934. The preparation of heparin from beef lung. *Tr. Roy. Soc. Can., Sec. V* 28:55-58.
15. ——— 1936. Studies on heparin. IV. Observations on the chemistry of heparin. *Biochem. Jour.* 30:1927-33.

## HEPARIN AND THROMBOSIS

16. Howell, W. H. 1928. The purification of heparin and its chemical and physiological reactions. *Bull. Johns Hopkins Hosp.* 42:199-206.
17. Jorpes, E. 1935. Das Heparin — eine Chondroitinpolyschwefelsäure. *Naturwiss.* 23:196-97.
18. Jorpes, E. 1936. On heparin, its chemical nature and properties. *Acta med. Scand.* 88:427-33.
19. Jorpes, E. 1936. Der Aminozucker des Heparins. *Zeitschr. f. physiol. Chem.* 244:253-56.
20. Charles, A. F., and Scott, D. A. 1933. Studies on heparin. II. Heparin in various tissues. *Jour. Biol. Chem.* 102:431-35.
21. Jorpes, E., Holmgren, H., and Wilander, O. 1937. Ueber das Vorkommen von Heparin in den Gefässwänden und in den Augen. *Zeitschr. f. mikro-anat. Forsch.* 42:279-301.
22. Staemmler, M. 1921. Untersuchung über Vorkommen und Bedeutung der histiogenen Mastzellen im menschlichen Körper unter normalen und pathologischen Verhältnissen. *Frankfurt. Zeitschr. f. Path.* 25:391-435.
23. Lison, L. 1935. La Signification histochemique de la métachromasie. *Compt. rend. de soc. de biol.* 118:821-24.
24. Bergström, S. 1936. Ueber Polysaccharidesterschwefelsäuren mit Heparinwirkung. *Zeitschr. f. physiol. Chem.* 238:163-68.
25. Holmgren, H., and Wilander, O. 1937. Beitrag zur Kenntnis der Chemie und Funktion der Ehrlichschen Mastzellen. *Zeitschr. f. mikro-anat. Forsch.* 42:242-78.
26. Wilander, O. 1938. Studien über Heparin. *Skand. Arch. f. Physiol. Suppl.* 15.
27. Biedl, A., and Kraus, R. 1909. Experimentelle Studien über Anaphylaxie. *Wien. klin. Wochenschr.* 22:363-70.
28. Eagle, H., Johnston, C. G., and Ravdin, I. S. 1937. On the prolonged coagulation time subsequent to anaphylactic shock. *Bull. Johns Hopkins Hosp.* 60:428-38.
29. Waters, E. T., Markowitz, J., and Jaques, L. B. 1938. Anaphylaxis in the liverless dog and observations on the anticoagulant of anaphylactic shock. *Science* 87:582-83.
30. Chargaff, E., and Olson, K. B. 1937. Studies on the chemistry of blood coagulation. VI. Studies on the action of heparin and other anticoagulants. The influence of protamine on the anticoagulant effect in vivo. *Jour. Biol. Chem.* 122:153-67.
31. Jaques, L. B., Charles, A. F., and Best, C. H. 1938. The administration of heparin. *Acta med. Scand. Suppl.* 90:190-207.
32. Chargaff, E. 1938. Studies on the chemistry of blood coagulation. VII. Protamines and blood clotting. *Jour. Biol. Chem.* 125:671-76.
33. Howell, W. H. 1925. The purification of heparin and its presence in blood. *Am. Jour. Physiol.* 71:553-62.
34. Quick, A. J. 1936. On the coagulation defect in peptone shock. A consideration of antithrombins. *Am. Jour. Physiol.* 116:535-42.
35. Best, C. H., Charles, A., and Cowan, C. 1937. The administration of heparin. *Am. Jour. Physiol.* 119:272-73.

## CHARLES H. BEST

36. Shionoya, T. 1927. Studies in experimental extracorporeal thrombosis. III. Effects of certain anticoagulants (heparin and hirudin) on extracorporeal thrombosis and on the mechanism of thrombus formation. *Jour. Exper. Med.* 46:19-26.
37. Murray, D. W. G., Jaques, L. B., Perrett, T. S., and Best, C. H. 1937. Heparin and the thrombosis of veins following injury. *Surgery* 2:163-87.
38. Best, C. H., Cowan, C., and MacLean, D. L. 1938. Heparin and the formation of white thrombi. *Jour. Physiol.* 92:20-31.
39. Solandt, D. Y., and Best, C. H. 1938. Heparin and coronary thrombosis in experimental animals. *Lancet* 2:130-32.
40. Solandt, D. Y., Nassim, R., and Best, C. H. 1939. Production and prevention of cardiac mural thrombosis in dogs. *Lancet* 2:592-95.
41. Murray, G. D. W., and Best, C. H. 1938. The use of heparin in thrombosis. *Ann. Surg.* 108:163-73.
42. Crafoord, C. 1937. Preliminary report on post-operative treatment with heparin as a preventive of thrombosis. *Acta chir. Scand.* 79:407-26.
43. Holmin, N., and Ploman, K. G. 1938. Thrombosis of central vein of retina treated with heparin. *Lancet* 1:664-65.
44. Ploman, K. G. 1938. Heparin treatment of thrombosis in central vein of retina. *Acta ophth.* 16:502-12.
45. Friedman, M., Hamburger, W. W., and Katz, L. N. 1939. Use of heparin in subacute bacterial endocarditis. *Jour. Am. Med. Assn.* 113:1702-03.
46. Kelson, S. R., and White, P. D. 1939. A new method of treatment of subacute bacterial endocarditis using sulfapyridine and heparin in combination: preliminary report. *Jour. Am. Med. Assn.* 113:1700-02.

**Part III. Some Aspects of Immunity  
and Chemotherapy**





# RECENT CHEMICAL TRENDS IN THE STUDY OF IMMUNITY

BY

Michael Heidelberger

*Associate Professor of Biochemistry, College of  
Physicians and Surgeons, Columbia University, and  
Chemist to the Presbyterian Hospital, New York*

I AM deeply appreciative of the honor of participation in your anniversary celebration, and am glad, too, that we are celebrating a fiftieth anniversary and not a hundredth, for the scientific study of immunity, whose chemical aspects I wish to discuss, is entirely a development of the last fifty years, just as is this Medical School. At the time of your founding, the connection between pneumonia and the pneumococcus had just been established. Sternberg, our early American bacteriologist, had conceived what was later experimentally founded as the Metchnikoff Theory, and must have been planning his authoritative and weighty (I say this feelingly) *Manual of Bacteriology*. Fifty years ago the first studies on toxins and antitoxins were under way, and there followed with great rapidity the epoch-making researches of Buchner, von Behring, Ehrlich, Bordet, and the host of other students of immunity who contributed to the firm foundations of serology and immunology.

When a foreign protein, called an *antigen*, is injected into an animal there usually follows, some time later, the appearance of new substances in the animal's blood serum. These substances, called *antibodies*, are characterized by their property of reacting with the antigen injected. When this interaction results in the formation of a precipitate it is called a *precipitin* reaction. Antigen-antibody interaction is specific; that is, antibodies arising from the stimulus of a given antigen, such as crystalline egg albumin, do not precipitate solutions of another albumin, such as crystalline horse serum albumin. More closely related proteins may, however, cause animals to produce anti-

## MICHAEL HEIDELBERGER

bodies which overlap in part, and reactions between such antigens and antibodies evoked by closely related antigens are called *cross* reactions.

An invading microorganism may be considered immunologically as a collection of antigens. If these antigens stimulate the production of antibodies these may interact with the invader, causing its direct or indirect destruction; and the animal is said to be immune. Antibodies to a microorganism may react not only with antigens extracted from the invading cell to give precipitin reactions, but may react with such components on the effective bacterial surface. The bacteria then clump together, and this special case of the precipitin reaction at the reactive microbial surfaces is called an *agglutinin* reaction.

Before proceeding to the discussion of theories of antigen-antibody, or immune reactions, let us consider briefly what has been learned recently about the chemical basis of the antigenic function and the chemical nature of antibodies.

Some of the factors affecting the antigenic behavior of proteins—and most antigens are proteins—are now fairly well understood. After the early indications of Obermayer and Pick, the researches of Landsteiner and his collaborators showed that introduction of halogen, nitro-, or diazotized aromatic amino groups into the aromatic or other cyclic nuclei of proteins caused the appearance of new serological specificities differing from the species specificity of the protein originally used. In the sera of a varying, often small, proportion of rabbits injected with these modified proteins these changes in specificity were so characteristic of the chemical grouping introduced (termed *haptens* by Landsteiner) as to overshadow the original species specificity and cause cross reactions between otherwise unrelated proteins. For instance, atoxyl-azo-chicken serum gave precipitates with the sera of some of the animals injected with atoxyl-azo-horse serum, although chicken serum itself does not precipitate antisera to horse serum. This reaction was found to be exceedingly sensitive, and in many instances position isomers and stereo-isomers (as also shown by Avery and Goebel in

## CHEMICAL TRENDS IN IMMUNITY

the case of sugars) could readily be distinguished. It was also found by Landsteiner and by Hopkins and Wormald that modification of the free amino groups of proteins by alkylation, acylation, or introduction of the phenylurethane grouping gave rise to new specificities, so that tyrosine, histidine, and possibly the other cyclic aminoacids are not the only portions of the protein molecule concerned with immunological specificity. Indeed it would seem that almost any chemical alteration at any portion of the molecule might produce a more or less characteristic change in specificity. Thus denaturation results in a partial alteration in serological reactivity, as does also reduction, presumably of  $-SS-$  linkages to  $-SH$ . Unpublished work in our laboratory by Davis indicates that phosphorylation, under conditions considered to affect only free aliphatic hydroxyl groups, may produce a radical change in specificity in a crystalline protein such as egg albumin.

As far as the natural proteins are concerned, questions of molecular shape and the arrangement of the component amino acids, whether or not these may ultimately be capable of expression in physical terms, are of the utmost importance. The immunochemist is therefore vitally concerned in the outcome of the current studies of Astbury, Bergmann, Felix, Langmuir, Waldschmidt-Leitz, and Wrinch.

The protein antigens which are directly involved in infectious disease and the manifestations of immunity to such disease are still among the least known and most poorly characterized. Many pathogenic microorganisms not only build up proteins as an integral part of their cell protoplasm but also secrete into the culture medium highly poisonous products, or toxins. Toxins are strongly antigenic, and the antibodies to which they give rise, the antitoxins, are of consistent therapeutic efficacy. Owing to the paramount importance of toxins in many diseases numerous attempts have been made to isolate them, but these have succeeded only recently. As a result of three independent investigations diphtheria toxin appears to be an easily denatured protein of 72,000 molecular weight. Its

## MICHAEL HEIDELBERGER

isoelectric point is at about  $pH$  4.1, and about 0.00045 mg. corresponds to one flocculation unit with antibody ( $I_t$ ).  $[\alpha]_D$  lies between  $-40^\circ$  and  $-50^\circ$ . About 14 per cent of the total nitrogen reacts as amino nitrogen, and about one-third of this disappears in the reaction with formaldehyde to yield anatoxin. Correlation of the toxicity with chemical constitution has not yet been possible.

Also vaguely characterized are the proteins of the bacterial cell. The classical method for their preparation involved extraction of bacteria with weak alkali and precipitation of the extracted protein with acetic acid. It was recognized that the product obtained in this way showed some of the properties of nucleoproteins and of mucoproteins. Recently, however, it has been possible to make a tentative and imperfect separation of the complex mixture of antigens in the bacterial cell, carrying out all operations in the cold in order to minimize the action of the active enzymes known to be present. By means of a preliminary acidification the mixture of protein salts in the cell is brought to the isoelectric point, or below. An extraction with buffer at  $pH$  6.5 then removes the most strongly acid fraction or fractions of the cell proteins, while subsequent successive extractions with increasingly alkaline solutions remove protein fractions of lower acidity. In this way a number of fractions have been obtained from the hemolytic streptococcus and the tubercle bacillus, and the products have been further fractionated by partial precipitation with sodium or ammonium sulfate. In each group several of the fractions showed immunological and chemical properties different from those of other fractions.

In the large group of encapsulated microorganisms to which the pneumococcus belongs, the dominant cellular antigen occurs in the capsular layer. The complete antigen has not yet been isolated, but each serological type of pneumococcus (and there are some forty) appears to be characterized by a chemically distinct polysaccharide to which these type-specific serological reactions are due. When isolated, the carbohydrates are

## CHEMICAL TRENDS IN IMMUNITY

found capable of entering into precipitin reactions with antisera from animals injected with the same pneumococcus type, and these antisera also agglutinate pneumococci of the homologous type, owing to interaction with specific polysaccharide on the bacterial surfaces.

In view of the importance of polysaccharides in determining the type specificity of encapsulated microorganisms and a definite fraction of the group specificity of most others it is surprising that the carbohydrate associated with many proteins has never been shown to react immunologically. Although drastic chemical treatment is often necessary for the separation of the apparently combined carbohydrate, the products obtained by the action of proteolytic enzymes have also failed to react with antiserum to the protein. Studies of this type have been made with egg albumin, with pseudoglobulin, and with thyroglobulin. In the instances in which the products of hydrolysis of carbohydrates from proteins have been isolated, the same three sugars — mannose, glucosamine, and possibly galactose — have been found. While this need not indicate the identity of the parent carbohydrates, it is possible that the same polysaccharide is present in most proteins. At the moment, therefore, it would seem that the carbohydrate in proteins, if actually chemically combined, as appears probable, either makes no contribution to the immunological specificity of the molecule as a whole, or else the contribution is not of the hapten type, but resembles that of certain azo substituents we have studied. Although they modify the original specificity these substituents may fail to stimulate the production of antihapten in the sense used by Landsteiner. If such an alternative applies, the presence of even the same carbohydrate could modify the specificity of proteins containing it to a greater or less degree without eliciting the production of antibodies capable of reacting with the isolated carbohydrate.

A somewhat similar state of affairs obtains within the large group of hemoglobins, in which the blood pigments of different species are known to have different specificities. As far as could

## MICHAEL HEIDELBERGER

be determined by qualitative tests, variations within the common prosthetic group, heme, such as substitution of CO for O, have no influence on the specificity. Moreover, undenatured globin is antigenic, and gives rise to antisera which do not permit a distinction between globin and the corresponding hemoglobin, at least when tested qualitatively.

After this brief survey of antigens let us turn our attention to antibodies. There is much evidence that these may in part be firmly fixed to cells. Most chemical studies on antibodies have, however, been carried out with those circulating in the blood, and I shall limit my discussion to these. It has long been known that circulating antibodies occur in one or more of the globulin fractions of serum. Partly as a result of the studies of Felton on the purification of antibodies in antipneumococcus sera and partly as a consequence of the quantitative studies from our own laboratory which I shall discuss it is now generally conceded that antibodies are not merely hypothetical substances of unknown nature associated with globulin, but are actually modified serum globulins.

If the protein nature of antibodies be accepted, it is still necessary to account for their formation and to discover how antibodies differ from normal proteins. Evidence is overwhelmingly against the old Buchner hypothesis that antibody specificity is explained by the actual entrance of antigen or antigen fragments into the antibody complex. Although unsupported by direct experimental evidence the theory put forward by Breinl and Haurowitz and later by Mudd seems more reasonable to the chemist. According to this the synthesis of normal serum globulin may occur in such a way that the spatial configuration of the cellular protoplasm impresses upon the globulin the spatial, chemical, and species-specific properties characteristic for the animal in question. Penetration of injected antigen or its partial degradation products to the sites of this synthesis could disturb the spatial relations which normally exist and distort them. This distortion might reasonably occur in a manner characteristic of the foreign material, so that any new globulin

## CHEMICAL TRENDS IN IMMUNITY

synthesized in this distorted manner would bear a certain spatial relation to the antigen. Hence if the new globulin molecules would again encounter the antigen in the circulation, or in vitro, interaction might be possible.

This theory, although unproved, does more than provide a chemically reasonable mechanism for antibody formation. It also accounts for the chemical similarity of antibodies to normal globulins, for differences between the two would be mainly those of configuration or arrangement of the component amino-acids. Recent studies by Dr. Florence R. Sabin with our red antigenic dye, R salt-azobiphenylazo-egg albumin, are not only in accord with this theory but extend it by pointing out the probable site of normal globulin synthesis as well as that of antibody. Dr. Sabin infers from her evidence that the phagocytic cells of the liver, spleen, tissues, and lymph nodes normally produce serum globulin, and that antibody globulin represents the synthesis of a new kind of protein by the same cells under the influence of an antigen. These pioneer studies should suggest a direct experimental approach to settle this vexing and difficult problem, so vital to the understanding of the physiological basis of immunity.

The early theory of Ehrlich that immune reactions such as the precipitin reaction involved the chemical combination of antigen and antibody in stoichiometrical proportions was opposed by that of Bordet, which held them to be merely adsorption phenomena. It was soon necessary to modify the latter view in order to account for the specificity of immune reactions, and an initial chemical reaction was gradually assumed, followed, in the precipitin and agglutinin reactions, by flocculation due to the presence of electrolytes. Arrhenius and Madsen stressed the reversibility of some antigen-antibody reactions, applied the law of mass action, and found analogies with the reaction of weak acids and weak bases. At the time these theories were proposed, little was known of the chemical nature of either antigens or antibodies, and the only analytical methods available were either of biological nature or the essentially quali-

## MICHAEL HEIDELBERGER

tative and entirely relative serological dilution methods, with their large capacities for error.

Before undertaking studies on the actual mechanism of the precipitin reaction, Dr. Kendall and I therefore found it necessary to abandon such techniques and to devise absolute micro-methods for the estimation of antigen and antibody, conforming to the criteria of analytical chemistry and yielding accurate data regardless of the presence of nonspecific protein. Precipitates formed from accurately measured quantities of antiserum and antigen dilutions were washed under definite conditions with 0.9 per cent sodium chloride solution and analyzed for nitrogen. In the proper regions of the precipitin reaction range this procedure permits the estimation of antigen or antibody in milligrams per milliliter.

The first instance of the precipitin reaction studied in this way was that between the nitrogen-free specific polysaccharide of type III pneumococcus\* and the partially purified antibody in a Felton solution prepared from type III antipneumococcus horse serum. This choice of "antigen" greatly simplified the analytical problem of the composition of the precipitates, for the polysaccharide made no contribution to the nitrogen precipitated. Once it had been shown that nonspecific nitrogen was not involved, the nitrogen estimations became a direct measure of the amount of antibody precipitated.

If a very small amount of S III is added to a relatively large amount of antibody (A), it is found that more than 40 mg. of antibody nitrogen may be precipitated for each milligram of S III. If increasing amounts of S III are added in separate portions to the same quantity of A, this ratio in the precipitate decreases steadily, with no evidence of discontinuity. In this region of the reaction range no S III can be found in the supernatants by the delicate serological test sensitive to a dilution of 1:10,000,000, so that it is assumed that all of the S III added is in the precipitate. In this region antibody is still in excess, as may be shown by addition of a little S III to a portion of the

\* Subsequently referred to as S III.



## CHEMICAL TRENDS IN IMMUNITY

supernatant. With still larger amounts of S III a region of the reaction range follows in which neither S III nor A is demonstrable in the supernatant from the precipitate, and this we have termed the *equivalence zone*. With still larger amounts of S III the polysaccharide finally appears in the supernatant, and in this region precipitation of antibody is at a maximum. When still greater quantities of S III are added, antibody nitrogen precipitated remains constant while more of the S III added enters into combination. Finally, in some antibody solutions, at least, the precipitate attains constant composition, and with larger amounts of S III, less and less precipitate is formed, until finally precipitation is entirely inhibited. The region of diminishing precipitation is therefore termed the *inhibition zone*.

In a sense these reactions are reversible, for the precipitate formed in the region of excess antibody takes up S III when shaken with a solution of the polysaccharide and even dissolves in relatively concentrated S III solutions.

The reversible shift in composition of the hemocyanin precipitate in either direction with antigen or antibody had been shown long before by von Dungern. It therefore seemed reasonable to postulate the following equilibria in the four limiting regions of the reaction range: at extreme antibody excess,  $S + 4A \rightleftharpoons SA_4$ ; at the midpoint of the equivalence zone,  $S + A \rightleftharpoons SA$ ; in the antigen excess region,  $SA + (x - 2)S \rightleftharpoons S_{x-1}A$ ; and in the inhibition zone,  $S_{x-1}A + S \rightleftharpoons S_xA$ . The italicized formulas represent precipitates, and in all formulas the composition is expressed in arbitrary units, not molecules. In the first two equations equilibrium must lie far to the right, as measurable dissociation could not be detected. It was shown that  $S_xA$  contained one more molecule of S than the precipitate with which it was in apparent equilibrium, confirming the belief of Müller, von Dungern, and Arrhenius in a soluble antigen-antibody compound in this zone, rather than the obscure "peptization" of the precipitate, which had been proposed by advocates of the colloidal theory despite Müller's direct chemical evidence to the contrary.

## MICHAEL HEIDELBERGER

Since S III is the highly ionized salt of a polymeric aldobi-  
onic acid, and antibody globulin, dissolved in physiological me-  
dia, probably exists as an ionized sodium chloride complex, the  
initial reactions, at least, may be ionic. The application of the  
mass law in some form would seem justified. The precipitin  
reaction between S III and homologous antibodies would then  
be merely a complex instance of a specific precipitation such  
as that between barium and sulfate ions or silver and cyanide  
ions. Even the inhibition zone would have at least a partial  
analogy in the well-known solubility of silver cyanide in excess  
cyanide solution.

In qualitative terms this interpretation of the precipitin re-  
action appears satisfactory, but difficulties arise in the quanti-  
tative formulation of the reaction in terms of the law of mass  
action.

With the aid of several assumptions, however, it was found  
possible to derive from the law of mass action a relation which  
accounts quantitatively for the S III-antibody reaction and  
many other instances of the precipitin reaction as well. These  
assumptions and simplifications are:

1. S III and antibody (A) are chemically and immunologi-  
cally multivalent with respect to each other; that is, each sub-  
stance possesses two or more groupings capable of reacting with  
the other.

2. Although the anticarbohydrate is known to be a mixture  
of antibodies of different reactivities it may be treated mathe-  
matically as if its average behavior were that of a single sub-  
stance, A.

3. For convenience of calculation the S III-antibody reac-  
tion is considered as a series of successive bimolecular reactions  
which take place before precipitation occurs.

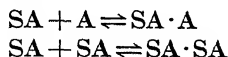
4. The mass law applies, so that the rates of formation of  
the reaction products are proportional to the concentrations of  
the reacting substances.

The reactions postulated are, in arbitrary units:

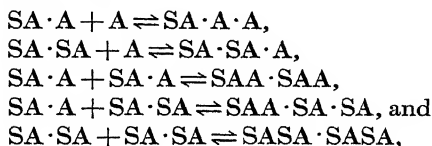


## CHEMICAL TRENDS IN IMMUNITY

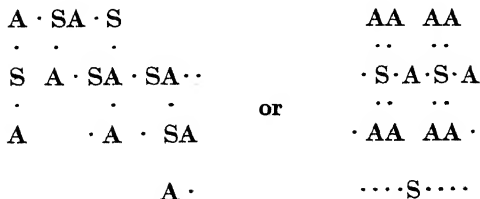
followed, for example in the region of excess antibody, by the competing bimolecular reactions due to the mutual multivalence of the components:



A third step would follow, in which the competing biomolecular reactions would be:



in which the first two reactions would occur only in the presence of enough A to carry the composition of the reaction product beyond the  $\text{SA}_2$  stage. Similarly, each compound formed in the third step would react with each other compound, or with more A, if present, to form still more complex substances; and the reaction would continue until particles would be formed large enough to settle from the solution. Precipitation would take place at this point, doubtless facilitated by the mutual discharge, with loss of affinity for water, of ionized or polar groupings brought together by the series of chemical reactions. The final precipitate would in each case consist of antibody molecules held together in three dimensions by S III molecules,



The process of aggregation as well as the initial hapten-antibody

# MICHAEL HEIDELBERGER

combination is considered to be a chemical reaction between definite molecular groupings.

The mathematical treatment of these reactions leads to equations of the form:

$$\text{mg. of antibody N precipitated} = 2RS - \frac{R^2S^2}{A}$$

in which A = milligrams of antibody N precipitated at a reference point in the equivalence zone and R = the ratio of A to milligrams of S III or antigen N precipitated at the same reference point. Both of the constants used therefore have chemical and immunological significance.

If both sides of the equation be divided by S, the resulting equation,

$$\frac{\text{mg. antibody N precipitated}}{\text{mg. S precipitated}} = 2R - \frac{R^2}{A} S$$

is that of a straight line. This linear relationship makes it possible to characterize an unknown type III antipneumococcus serum or antibody solution in the region of excess antibody by two or, better, three analyses, in duplicate. If the ratio of antibody N to S III precipitated be determined for two or three different amounts of S III in the region of excess antibody and a straight line be drawn through the points so obtained, the

intercept on the y axis =  $2R$  and the slope =  $-\frac{R^2}{A}$ . With the R and A values at the beginning of the equivalence zone calculated in this way the amount of antibody nitrogen precipitated by any quantity of S III less than  $\frac{A}{R}$  may be calculated with a fair degree of accuracy.

Despite the wide variation in the behavior of individual sera these and similar expressions for the other ranges of the reaction permit the complete description of the precipitin reaction between S III and an unknown antiserum without an unduly burdensome number of microanalyses or the sacrifice of a large amount of material.

## CHEMICAL TRENDS IN IMMUNITY

Application of these principles to precipitin systems in which the antigen is a protein is complicated by the necessity of distinguishing between antigen and antibody nitrogen if the composition of the precipitate is to be determined. This was accomplished with the aid of the red protein dye already referred to. This was freed from fractions reactive in anti-egg albumin sera and then injected into rabbits. In this antigen-antibody system antigen could be estimated colorimetrically in an alkaline solution of the washed specific precipitate, while antibody was determined by difference after a total nitrogen estimation. The equations given were found applicable in this system as well. With the aid of the information gained, a study was made of a colorless protein, crystalline egg albumin, and its homologous antibodies; and this instance of the precipitin reaction was also found to be quantitatively described over a large part of the reaction range by the theory. The precipitin systems with crystalline horse serum albumin and various thyroglobulins and their homologous antibodies were also found to conform.

It is also possible, as Kabat and I have shown, to adapt the microanalytical methods used for studying the precipitin reaction to the estimation of pneumococcus agglutinins and with these methods to show that type I pneumococcus and its homologous antibody undergo reaction according to the same type of equation. The reaction is, however, simpler, since the exigencies imposed by the reactive bacterial surfaces limit the range of combining proportions of bacterial polysaccharide and antibody.

As might have been anticipated from the view that specific bacterial agglutination differs from specific precipitation only in that the former reaction takes place on particulate matter, the latter between two dissolved reagents, type-specific pneumococcal anticarbohydrate was found to be quantitatively the same whether measured as agglutinin or as precipitin. It was also found that the entire course of these instances of specific bacterial agglutination could be accounted for, as in the precipitin

## MICHAEL HEIDELBERGER

reaction, on the basis of a chemical reaction between multivalent antibody and multivalent antigen, without assumptions as to electrical potential or cohesive force.

A very recent extension of the quantitative agglutinin method has been to rabbit antisera to *H. influenzae*, type B. The knowledge of the antibody content of these sera was made urgent by indications obtained by Dr. Alexander at the Babies Hospital, New York, that such sera may produce a therapeutic effect in influenzal meningitis in children. As a result of improved methods of immunization of the rabbits and analytical control of the antisera there are now available sera and purified antibody solutions of five to ten times the actual antibody content of the best available at the inception of the work.

The above microanalytical methods for the study of the precipitin and agglutinin reactions have met with general acceptance and are coming into increasing use for the standardization of antipneumococcus sera in place of the expensive, inaccurate, and time-consuming mouse protection test. The quantitative theory based on data acquired by these methods is still in the controversial stage, but regardless of its ultimate fate it has already had several far-reaching consequences and has permitted some predictions that might not otherwise have been made.

For example, a study of the effect of strong salt solutions on the reaction between pneumococcus polysaccharide and homologous antibodies showed that the diminished precipitation and the decrease in the values of both constants in the equations was not due to increased solubility of the precipitate. On the basis of the quantitative theory a reversible shift in the equilibrium between polysaccharide and antibody was indicated, and the prediction was made that use could be made of this shift for the isolation of pure antibody. If, for example, 0.1 mg. of S III precipitated 1.24 mg. of antibody nitrogen from 1 ml. of a given antiserum in 0.15 M sodium chloride and only 1.01 mg. of nitrogen in 1.75 M sodium chloride, it should be possible, if the equilibria involved were reversible, to add 1.75 M sodium

## CHEMICAL TRENDS IN IMMUNITY

chloride to the washed precipitate formed in 0.15 M sodium chloride and dissociate 0.23 mg. of nitrogen, representing analytically pure antibody. Practical difficulties in washing and handling larger amounts of material prevented the immediate realization of this ideal, but solutions of 90 to 98 per cent purity were readily obtainable in a single step from many antipneumococcus horse and rabbit sera of various types. With refinements in technique, analytically pure antibody globulin was isolated.

Some of the highly purified antibody globulin solutions were studied in Upsala, Sweden, in the Svedberg ultracentrifuge. Pedersen and I found that the type III pneumococcus antibody produced in the rabbit showed only a single component with a sedimentation constant of  $7 \times 10^{-13}$ , the same as that of the principal component of normal globulin. A similar preparation from horse serum was also monodisperse, but showed a sedimentation constant of  $18 \times 10^{-13}$ , close to that of a minor globulin fraction of relatively high molecular weight present in all mammalian sera. Removal of the cell contents at different levels after centrifugation showed that the antibody had actually sedimented with the heavy portion. This pointed to entirely different mechanisms for the production of pneumococcus anticarbohydrate in the horse and in the rabbit. The same antibodies produced in other species of animals were purified in the laboratory of the Presbyterian Hospital and were shown by Kabat and Pedersen to exhibit similar differences.

From these studies it is clear that pneumococcus anticarbohydrate and anti-egg albumin engendered in the rabbit have the same molecular size as that of the principal component of normal serum globulin, while the former antibody, as produced by the horse, cow, and pig, has a molecular weight six times as great, corresponding to a minor component also present in all normal sera. If there could any longer be doubt as to the actual protein nature of antibodies, it would be dispelled by the typical protein electrophoresis curves found for several of the above antibody solutions by Tiselius and by the study of the electro-

## MICHAEL HEIDELBERGER

phoretic behavior of antibodies in whole antisera made by Tiselius and Kabat.

The new information gained from this work has had its practical repercussions as well, for it has furnished much of the theoretical background and practical methods of control for the use of antipneumococcus rabbit sera in the treatment of pneumonia.

One other feature of the ultracentrifugal sedimentation diagrams is of interest. Evidence has already been given for the existence of actual chemical compounds in the inhibition zone of the SIII-antibody system, but the visualization of such compounds in the ultracentrifuge would seem to place their existence beyond question. It should be remembered that the specific precipitate was in this instance composed of aggregates of egg albumin, with a sedimentation constant of about  $3.6 \times 10^{-13}$ , and antibody with  $s = 7 \times 10^{-13}$ . The presence in the solution of a small number of components of higher sedimentation constant than either would seem to indicate the formation of definite compounds, for if the precipitate were merely "peptized" by the excess of antigen, as the colloidal explanation goes, there is no a priori reason why this should stop short of the actual reaction components themselves.

An additional result of the quantitative precipitin and ultracentrifugal studies is the possibility, now that the molecular weights of the reactants are known, of writing empirical formulas for the composition of specific precipitates at limiting points in the reaction range. The formulas are to be considered as average compositions, and not those of single molecular species. However, if the conception of precipitate formation by the union of multivalent antigen with multivalent antibody is correct, each particle of precipitate is a molecule of vast size, of approximately the empirical composition indicated for a given point in the reaction range. Nor are the compositions given (ranging roughly from  $\text{Antg}(\text{Antb})_6$  to  $\text{Antg Antb}$ ) of such extreme range as to render preposterous a treatment of the reaction according to the laws of classical chemistry.

Another prediction, made from the quantitative agglutinin



## CHEMICAL TRENDS IN IMMUNITY

theory, explains the role of salts in these immune reactions on a basis different from that currently held. The reversibility of the precipitin reaction, in the sense that a precipitate may be shifted from one region of the reaction range to another by addition of either component, suggested a similar possibility for the closely related agglutinin reaction. It was predicted that if, for example, type I pneumococci were agglutinated with a large excess of antibody, and that if the excess of antibody were removed by thorough washing and the agglutinated pneumococci were resuspended in physiological salt solution, they would re-agglutinate immediately into larger clumps on addition of appropriate amounts of fresh type I pneumococci or type I specific polysaccharide. It was also predicted that no change would occur on addition of, for example, type II pneumococci or type II specific polysaccharide, although conditions of electrical potential and cohesive force would be nearly identical. These predictions were fully verified. The experiments were considered to indicate that in this way the process of specific bacterial agglutination had been, in a sense, interrupted at an early stage (much antibody and few pneumococci), and that resumption of the agglutination process under controlled conditions showed that aggregation into large clumps was merely a continuation of the chemical union of multivalent antigen on the effective bacterial surface with multivalent antibody on other effective bacterial surfaces. The assumption of an initial static antigen-antibody union and the subsequent building up of large aggregates through a combination of cohesive force and salt-lowered potential are thus contrary to actual experimental data. The alternative explanation that dissociation of antibody could have occurred was eliminated by washing the initially agglutinated pneumococci until the supernatant no longer agglutinated added pneumococci, and also by the obviously greatly larger clumps formed in the reagglutination.

Not only, then, is there evidence that the process of specific bacterial agglutination (and hence the precipitin reaction) differs from other aggregating systems long used as analogies, in

## MICHAEL HEIDELBERGER

that it behaves as a single chemical reaction from beginning to end, but the same experiments place the function of salts in another light. The effect of salts may be interpreted as the purely secondary one of providing ions for the ionized salt complexes in which form antibody probably reacts, and, in addition, of minimizing electrostatic effects due to the presence of many ionized groupings on the particles, effects which might interfere with the completion of particulation by chemical interaction.

From this review of recent chemical trends in the study of immunity it is apparent that new organic chemical techniques, powerful physicochemical tools, the application of the rigorous quantitative micromethods of analytical chemistry, and the theories that have emerged from the use of such methods have advanced our understanding of the chemical basis of immune processes to a considerable degree. It is confidently to be expected that the continued cooperation of the organic chemist and the physical chemist with those in the fundamental medical sciences and in clinical medicine will result in a widening of our knowledge of immunity in the next fifty years to a far greater extent than in the period since the founding of this Medical School.

### *References*

With the exceptions added below, references to the original articles cited will be found in the following recent reviews by the author:

- 1938. Relation of proteins to immunity. Chapter XVII in Schmidt, C. L. A., *The Chemistry of Amino Acids and Proteins*. Thomas, Springfield, Ill.
- 1938. Protein constitution and immunological behavior. *Cold Spring Harbor Symposia on Quantitative Biol.* VI:369.
- 1938. Les Aspects chimiques de la fonction antigène et leurs rapports avec les agents infectieux. *Rev. d'immunol.* 4:293.
- 1939. Chemical aspects of the precipitin and agglutinin reactions. *Chem. Rev.* 24:323.
- 1939. Quantitative absolute methods in the study of antigen-antibody reactions. *Bacteriol. Rev.* 3:49.

In addition:

- Sabin, F. R. 1939. Cellular reactions to a dye-protein with a concept of the mechanism of antibody formation. *Jour. Exper. Med.* 70:67.
- Alexander, H. E., and Heidelberger, M. 1940. Agglutinin and precipitin content of antisera to haemophilus influenzae, type B. *Jour. Exper. Med.* 71:1.

# THE BIOLOGY OF ANIMAL VIRUSES

BY

Robert G. Green

*Professor of Bacteriology, University of Minnesota*

FOR a generation filterable viruses and the diseases caused by them have been a much discussed mystery of medical science. The ultramicrobes present a paradox. Inasmuch as they obviously possess the power of reproduction, a property known to be one of living matter only, they would appear to be alive. On the other hand, their smallness seems to preclude a living nature; most of them have been found to be of a size that would not support the complexity of a living organism. Shortly after the innovation of bacterial filters of such fine porosity that they retain ordinary bacteria, Iwanowski in 1894 proved that the disease of tobacco known as mosaic disease could be reproduced by inoculation of material that had passed through bacteria-proof filters. Four years later the same phenomenon was independently discovered by Beijerinck, who described the infective agent as a *contagium vivum fluidum*. An extension of this concept to diseases of the animal world was made by Loeffler and Frosch in 1898, when they showed that the virus of foot-and-mouth disease of cattle could be demonstrated in the fluid that passed through bacteria-proof filters. These findings were so contrary to the then recently propounded theory of microbial disease, and the phenomenon itself seemed so unaccountable, that further information along similar lines was accumulated slowly. During the past decade, however, the expansion of knowledge of filterable viruses and virus diseases has been rapid and has led to the beginnings of a new science. It would seem that the information now available on viruses and virus diseases is amenable to a systematic arrangement if approached from the viewpoint of parasitism. Although the work of Stanley, introduc-

ing the crystallization of viruses and ideas as to the possibility of their being chemical in nature, has recently given credence to the theory that viruses are chemical units rather than microbes, no facts nor concepts have been advanced that preclude the treatment of our knowledge of viruses in accordance with a biological system. A biological approach to the study of viruses allows an explanation of their origin and their nature. It clarifies the fundamental aspects of the diseases they produce and at the same time explains the characters that seem to place viruses in a chemical category.

*Concepts of disease.*—Men first believed that disease was a manifestation of possession of an individual with a demon, and consequently early human medicine concerned chiefly methods of casting out devils. This theory of disease persisted for many centuries and was discarded only after the discovery of microbes and the proof that they caused certain infections which spread from one person to another. Since the time of Koch and Pasteur whole sciences have developed that are based upon the discovery of microbes and their relationship to infection. Nevertheless our basic concept of disease has changed but little. Human beings are the center of our interest and we still interpret disease in terms of its effects on man. We have failed to recognize that microbes are species of life ordered by nature in reproduction and survival, and that their activities represent primarily the life history of biological species, representing disease only secondarily, from the viewpoint of animal or human welfare. For sixty years we have conducted detailed studies on microbes. We have learned to differentiate one species from another and have discovered that certain ones, upon invasion, cause specific symptoms. We have even carried on intensive research on chemical changes that occur as the microbes grow. However, we have given little attention to the activity of microbes as a means of perpetuating their species and to the effect of parasitic life upon the microbes themselves.

The biological concept of disease that I wish to present is one in which the microbe takes the center of the stage. The microbe

## THE BIOLOGY OF ANIMAL VIRUSES

is a species of life, endowed with the same vital mechanism that has given rise to our own being, and with all of the characters devised by nature for the perpetuation of a species. The activities of the microbe are in some way related to its well-being and to the preservation of its species. According to this concept, the body of man or other animals afflicted with disease is but an environment exploited by a species of microbic life.

As long as we dealt with infections caused by visible microbes, both satisfaction of scientific inquiry and cure of the disease could be achieved through study of the microbes as invading demons. In filterable viruses we have something far smaller than the microscopically visible germs or microbes. These ultramicroscopic incitants of disease are variously called filterable viruses, viruses, and ultramicroscopic microbes. Since the study of diseases caused by viruses cannot be based upon microscopic observation of the organism itself, the investigation must be founded entirely upon the effects produced by the ultramicrobes.

It is obvious from the size of viruses, indirectly determined, that no microscope can be built, in the present state of our knowledge, that will make these agents of disease directly visible. We follow their progress through a series of effects, and to understand a virus disease we must understand the significance of these effects. This requires a reversal of our usual approach to the problem of disease. We must, to a certain extent, consider the animal body simply an environment to which the invading microbe or ultramicrobe has adapted itself, and must concentrate upon learning and interpreting the life history of the microbic species.

*Relationship of pathogenic microbes to free-living species.—*

It is agreed that all species of parasite are related to similar, nonparasitic species that live free in nature, entirely independent of a host. Free-living species of microbe are ubiquitous; they are found in ponds, in stagnant waters, and in decaying material of all kinds, where they carry on important biological processes in the conversion of matter. Pathogenic species of

## ROBERT G. GREEN

microbe are accepted as races of microbes that have adapted themselves to the environment of an animal body. Certain of the pathogenic species have become so highly adapted that they are no longer able to carry on vital processes unless they are within a living animal body. Such species are termed obligate parasites. The preservation of species of this kind is contingent upon assured passage from one animal to another. The microbes must have specialized ability to obtain entrance into animal bodies, to reproduce there, and to conduct themselves in such a way that their progeny may find a means of escape from one animal body and gain entrance to another.

A great variety of adaptations, from simple to very complex, has been acquired by visible microbes to insure their continued existence in animal hosts. Typhoid bacilli grow in the wall of the gastro-intestinal tract and after reproduction escape from the body in large numbers. The malarial protozoon, on the other hand, does not grow in the human body at a point from which it can escape to a new host. It relies upon the bite of a particular species of mosquito for transference from the blood of one animal to the blood of another. In the mosquito this microbe undergoes a series of reproductive changes which facilitate its transfer from one host to another.

Considering a disease process from the standpoint of the microbe we see also that a natural limitation is placed upon the devastating effects of a disease. Any microbic species that destroys its host rapidly without permitting its own progeny to escape and find a new one is committing race suicide. The welfare of the organism requires that it allow as great an opportunity as possible for escaping and finding a new world to live in. The tubercle bacillus has developed one of the most favorable adaptations for the preservation of its kind. It grows almost exclusively in lung tissue, where it produces a slow, chronic, localized infection. When the lesion enlarges and breaks into the bronchial tree, escape to a new host becomes possible. Organisms are delivered into the outside world over a long period of time, so that ample opportunity is afforded for invasion of new

## THE BIOLOGY OF ANIMAL VIRUSES

hosts and perpetuation of the species. Occasionally tubercle bacilli stray from the lung tissue and grow throughout the animal body. This results in a quick infection of many organs and destruction of the environment, whereupon possibilities of further propagation of the species are lost. Tuberculosis in the past has been one of the greatest scourges of mankind because of the very successful adaptation of the tubercle bacillus to growth in the bodies of man and other animals.

When the problem of filterable virus diseases is attacked from the viewpoint of microbic adaptation, the mystery surrounding them is dissolved. On a former occasion I presented briefly a concept which sets forth the proposition that intracellular microbes are gradually transformed, by a series of adaptations, to incomplete forms of parasitic life now generally called filterable viruses. While this has been termed a theory of viruses, it is really an integral part of the concept of disease here presented. An understanding of viruses and virus diseases follows logically when an analysis of infectious diseases is made from the standpoint of the microbe itself.

*Similarity of microbic and virus diseases.*—There is no significant difference between the diseases caused by microbes and those caused by ultramicrobes. Typhoid fever is caused by a bacterium that is visible under a microscope and can be readily identified. It has been conclusively established that this microbe enters the human body by way of the mouth. For a period it grows throughout the body, but finally it is localized in the wall of the intestine, from which large numbers of its progeny pass to the exterior to gain access to another individual and there repeat the process. Virus diseases such as influenza, in which the ultramicrobe cannot be seen, follow the same sequence of events. A filterable virus enters an animal body and reproduces there; its offspring escape to cause infection in other individuals.

Similarly there is no basic difference between microbic and virus diseases, in which adaptations of the causative organisms are more complex. In malaria the visible parasite is securely locked within an animal host, for it does not grow in superficial

## ROBERT G. GREEN

tissues from which it might escape directly to another host. As stated before, this protozoon depends upon the bite of a mosquito for removal through the skin of the infected animal and implantation in the blood of another animal of a species in which it can grow. The virus of yellow fever, a typical filterable virus, is dependent upon transfer by a mosquito in exactly the same way. In both cases the agent of disease grows within the mosquito through a definite cycle before it reaches a position within this vector from which it can pass during the process of biting.

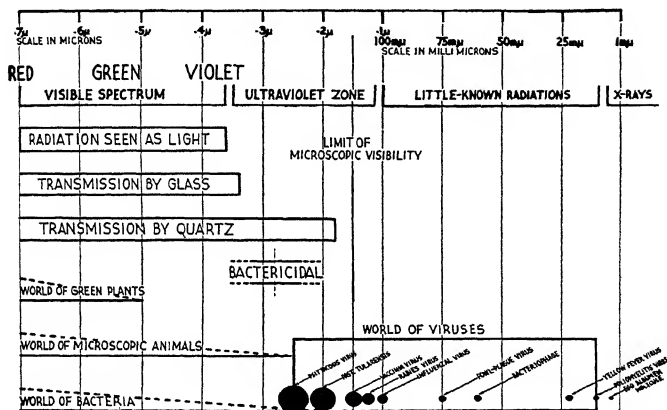
Thus we see that in important characters there is parallelism between microbic and ultramicrobic diseases. It seems reasonable, therefore, to regard virus diseases as biological processes and viruses as biological units comparable in many ways to visible microbes.

*Description of viruses.*—Filterable viruses may be described as pathogenic agents ranging from microbes scarcely discernible under the microscope to disease incitants that are particles about one-twenty-fifth the size of the smallest microbes. The viruses are differentiated from bacteria and protozoa, first, by their smaller magnitude and, second, by their inability to grow on artificial media. The smaller magnitudes of the viruses make them ultramicroscopic, or invisible under a microscope, and allow them to pass through specially constructed filters impervious to the larger bacteria. In size as well as in other characteristics, however, there is no sharp line of distinction between microbes and filterable viruses. The larger of the filterable viruses, such as the virus of parrot fever or psittacosis, can be seen by means of a microscope and are, in most respects, like the smallest of the bacteria and protozoa. Certain bacteria are classed as obligate parasites because they are unable to reproduce on artificial media unless some body fluid or tissue extract of the host is added to assist their growth. Some of the larger viruses, which are just visible under a microscope, seem to be similar to small bacteria except that their parasitism is more obligatory, so that they will not grow at all on an artificial



# THE BIOLOGY OF ANIMAL VIRUSES

medium. In some cases it is difficult to determine whether a small organism should be classified as a bacterium, a protozoon, or a filterable virus. *P. tularensis* is defined as a true bacterium. Like some viruses it is barely discernible under the microscope and will pass through filters that withhold most bacteria; but since it will grow on an artificial medium if the amino acid, cystine, is added, it is classified as a microbe. The formerly so-called virus of pleuropneumonia in cattle can be seen under the



**Figure 1. Magnitude relationships of viruses.**

microscope, is easily filtered, and grows on artificial media. It has therefore now been reclassified as a bacterium.

The various species of viruses present a series of sizes continuously decreasing from the magnitude of the smallest bacterium to approximately the size of a large protein molecule. In Figure 1 are shown the relative sizes of viruses in comparison with the lengths of radiations and the smallest of plants and animals. It is seen that the size of viruses ranges from that of radiations in the ultraviolet zone to the size of X-rays. The largest known virus is the psittacosis virus, which is greater in magnitude than *P. tularensis*, cited above as a true bacterium.

## ROBERT G. GREEN

The vaccinia virus, or smallpox virus, represents the tiniest particle that can be seen by means of a microscope. The viruses of rabies and influenza, which are slightly smaller than the vaccinia virus, are well beyond the limit of microscopic visibility. Two of the most minute viruses are those of yellow fever and infantile paralysis. The latter is only slightly larger than the molecule of egg albumen. It is also shown in Figure 1 that so far as sizes of organisms are concerned the world of microscopic animals overlaps that of the viruses, and the world of bacteria extends well into the world of viruses. Consequently there is no sharp demarcation between the viruses and the protozoa or microscopic animals, nor between the viruses and the bacteria or microscopic plants.

*Theories of viruses.*—A number of ideas and theories have been advanced about the nature of filterable viruses. Most of them, dealing with the smallest of the viruses, have been offered to explain how a particle can be as minute as a large molecule and still retain the property of reproduction. A filterable virus was first described by Beijerinck as a contagious fluid. D'Herelle, from his studies on bacteriophage, visualized a whole world of undiscovered, free-living, and parasitic ultramicroscopic forms of life. He called this group of organisms *Protista*. That this is an unreasonable view I shall point out in some detail.

It has been usual to think of viruses as a phenomenon distinct from all others and to try to discover an entirely new explanation of them. For example it has been conjectured and some experimental work has been done to demonstrate that viruses are similar to enzymes in that they originate spontaneously from the tissues of the animals and plants in which they are found as infectious principles. Rahn has suggested that a virus is a feral gene which has escaped from a cell and preys parasitically upon other cells. Thus far experimental work and philosophical dissertations have failed to substantiate the theory that viruses are derived from host cells or that they may logically be classed as intimately related to enzymes, as they have been in the past.

## THE BIOLOGY OF ANIMAL VIRUSES

Impetus has recently been given this view, however, by the crystallization of certain of the smaller viruses by Stanley. Work on crystallization has fostered the general belief that viruses are relatively simple chemical compounds. This supposition implies that they are something quite different from bacteria and protozoa and that virus infections are dissimilar to diseases produced by visible microbes. The fallacy here arises from forgetting that filterable viruses are a large and varied group and present a gradation in size from microscopic microbes to the exceedingly small forms which are subject to a process simulating crystallization.

Rather than attack the problem of viruses by first studying those of smallest magnitude, it seems reasonable to begin with a study of those of largest magnitude, which are so similar to microbes. A satisfactory explanation would, of course, have to encompass the entire group, both large and small. The nature of the larger viruses is readily explained by concepts that apply also to microscopically visible bacteria. A further application of this reasoning accounts for the continuous gradation in sizes of viruses and, finally, for the origin and nature of the smallest viruses, without recourse to an unnatural, a supernatural, or a revolutionary view of the nature of viruses and virus diseases. In fact the concept of the origin and nature of viruses becomes very ordinary in terms of biological parasitism.

*Magnitudes of living forms.*—Modern, intensive studies in microbiology over a period of fifty years have failed to reveal the existence of free-living microbes smaller than those of visible size. When an organic or inorganic solution undergoes changes that indicate the presence of living microbes, visible microbes can invariably be demonstrated by microscopic examination. The extensive use of bacteria-proof filters offers critical evidence that free-living ultramicrobes do not exist. Filtered solutions of organic material usually show no biological change indicative of microbial growth. When such a change does appear, it is possible to verify the presence of visible microbes, and the conclusion must be drawn that the filter used was defective and allowed

## ROBERT G. GREEN

microbes to pass. There is, of course, the special case of bacteria such as sulfur bacteria, which regularly penetrate bacteria-proof filters. Their passage is possible because the average size of the microbes is close to the limit of visibility. The smallest of the forms may be just below the limit of visibility and escape from filters that retain ordinary bacteria. Upon subsequent growth the small forms again give rise to forms that are within the limit of visibility.

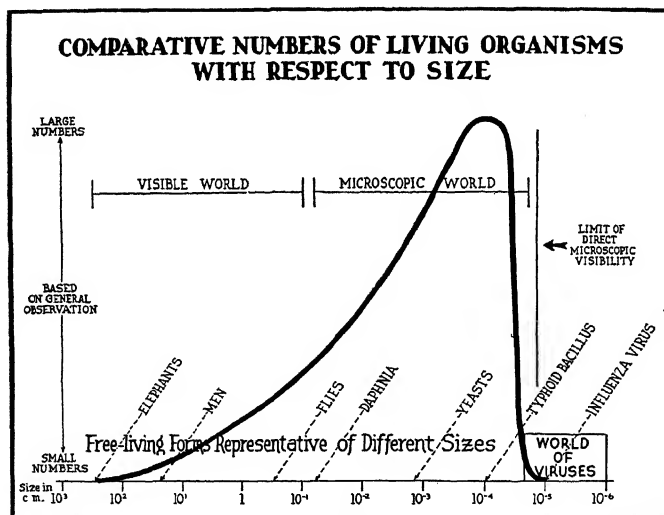


Figure 2. Comparative numbers of living organisms with respect to size.

A representation of the numbers of individuals and the numbers of species of living organisms, visible and ultramicroscopic, in relation to magnitudes, is given in Figure 2. The curve is based on the general observation that there are few large animals and innumerable small ones. Of the large animals such as buffaloes and elephants there are comparatively few species, and not many of these animals exist in the world at any one time. Species of animals of a lesser magnitude, say the size of

## THE BIOLOGY OF ANIMAL VIRUSES

mice, are much more numerous, and the individuals of each species more abundant. The numbers of species of organism of still smaller size, such as insects and tiny plants, are extremely great, and the numbers of individuals of such species multitudinous. In the microscopic world, the world of bacteria and protozoa, species become difficult to differentiate, and the numbers of individuals of each species are incomprehensibly great. The population of protozoa and bacteria in a single pond at any one time probably exceeds the number of all the elephants that have existed on this earth. But as we follow the magnitudes to still more minute size, we see an abrupt decrease in numbers. Beyond the limit of microscopic visibility we do not know of a single form of free-living life. It is probably significant that the lower limit of the magnitudes of free-living organisms corresponds with the limit of direct microscopic visibility, which approximates about half the average wave length of the visible spectrum. The ultramicroscopic forms of life that have been recognized are all pathogenic agents, which produce disease in animals and plants. It seems indubitable that all forms of life smaller than a given size are dependent for their existence upon free-living organisms larger than that size.

*Viruses as adaptive forms from the microbic world.*—We see the pathogenic bacteria as a few specialized species of a large group of free-living organisms. However, we cannot consider the ultramicrobes a specialized group developed from the world of free-living ultramicrobes, for there is no evidence that free-living ultramicrobes exist.

It is generally accepted that pathogenic forms have been derived from free-living forms. The bloodsucking insects have relatives that live on vegetable juices. Worms that are parasitic and produce disease have free-living relatives from which they obviously descended by retrograde processes. The pathogenic microbes and protozoa are clearly related to similar forms that live a free and self-sufficient existence. We infer, therefore, that viruses are related to free-living forms, from which they

## ROBERT G. GREEN

developed, or that they represent pathogenic particles which have had an entirely different origin.

A true picture of the nature and position of viruses in the living world is obtained from a consideration of adaptive change associated with parasitism, and from an application of the concept of such change to pathogenic microbes. Let us first inquire briefly into the adaptations of parasitism as we see them in the visible world.

One of the most outstanding parasitic adaptations has been made by the tapeworm. This flat worm, found as a parasite in man and many other animals, has become thoroughly adjusted to life within the gastro-intestinal tract. Bathed in digestive juices and surrounded by digested food, it has had no need of a digestive system of its own. Through the process of evolutionary change in that particular environment, the tapeworm has lost all vestiges of a gastro-intestinal tract. Among visible forms of life anatomical modifications are seen in all degrees, from slight changes to this very drastic loss of an entire digestive system. It is important to recognize that because of its simplification the tapeworm is an obligate parasite. Moreover it is important to realize that the tapeworm has become an incomplete or partial form of life. It is no longer able to make its own living in the ground alongside the earthworm. The process of modification so common in the visible world plainly occurs in two stages: first, the loss of a function; second, the loss of the anatomical structure necessary to carry out that function.

It is clear that comparable adaptations, also seen in many degrees, have accompanied the gradual change of certain free-living forms of bacteria into parasitic or pathogenic microbes. Free-living species of bacterium which live habitually in pond water are capable of growth throughout a wide range of temperatures, and for optimum growth require a relatively low temperature, corresponding more or less to the average for pond water. Similar but parasitic species of microbe show a marked conformity to their environment in respect to this physical factor and do not grow throughout so broad a range as do the free-

## THE BIOLOGY OF ANIMAL VIRUSES

living forms. The optimum temperature for growth of pathogenic bacteria corresponds closely to that of their host and is much higher than that of the free-living forms found in pond water. Among the various disease-producing microbes one may observe degrees of environmental dependencies varying from adaptations of temperatures to changes that prohibit their growth except in intimate association with a living host, and result in a very precarious existence of the microbes in the outside world.

It is evident that microbic life is subject to the same types of adaptation in the process of parasitism that are noted for the larger, free-living forms, which can be observed directly. Obligate parasitism by microorganisms must therefore be represented by a loss of functions, for which the parasite relies upon the host. Many of the pathogenic bacteria are dependent for growth on substances found in their hosts. Generally the nature of these compounds, or the purpose they serve, is unknown. In the case of *P. tularensis*, the organism that causes tularemia, we have an example of a known loss of function. *P. tularensis* has lost some faculty of the metabolism of cystine. This organism will not grow on ordinary bacterial media that support reproduction of the typhoid bacillus; but when the amino acid, cystine, is added to the media it grows as well as many saprophytic bacteria. It is not difficult to conceive extension of this retrograde process to a greater loss of the functions supplied by the host.

It is to be expected that loss of a function by a parasite will be followed by disappearance of the anatomical substance that performs the process. The losses would simplify the microbe and make it smaller, changes distinctly to the advantage of an intracellular microbe. The retrograde process, developing step by step, could stop at various stages in different species and thus produce a gradation in magnitudes of intracellular parasites. In certain cases the process might continue until extreme simplification resulted in a microbic fragment consisting of little more than the elements of reproduction and a few hapten

## ROBERT G. GREEN

groups, which would maintain the identity of the species. A partial microbe would be an incomplete form of life corresponding, in fundamental respects, to the tapeworm; it would be a complete, functioning individual only if immersed in the protoplasm of the host cell to which it had become adapted. Its own limited vital processes, complemented by those of the host cell, would allow it to grow and reproduce.

Depending upon its complexity, the incomplete form might or might not contain the spark of life. If relatively few vital processes were lost, the virus would still be a living unit, but one whose make-up was inadequate and whose vital functions must be complemented. A virus of this kind might die easily, as does the distemper virus. On the other hand, the retrogradation might continue to a point where only a residuum of the original microbe remained. The ultramicrobe would then no longer possess the spark of life, but would be dependent upon the host cell not only for most of its vital functions but also for the vital dynamics to energize its parasitic reproductive process.

The tremendous possibilities for adaptation within the protoplasm of a cell serving as host to an intracellular microbe appear to be the basis for the most extensive functional and anatomical changes that can be achieved through retrograde evolution. A microbe living in the gastro-intestinal tract has available only the limited chemical processes of digestion to which it may adapt itself. A microbe that has attained the ability to grow within tissue, but between cells, has a much wider range of chemical activities and vital processes to make use of. But here too possibilities for retrogression are much restricted. A microbe that has acquired the ability to penetrate to the interior of a cell, however, and to live there without greatly disturbing the metabolism of the cell, has almost unlimited opportunities for adaptation. It is bathed in all the chemical processes of protoplasm and may appropriate them for its own use. It would seem that when a microbe has adapted itself sufficiently to enter a host cell it has attained the greatest possibility for using the metabolic processes of the host to satisfy its parasitic tendency.



## THE BIOLOGY OF ANIMAL VIRUSES

The significant features of this concept of viruses are the capacity of microbes to invade the protoplasm of a cell, the ability to reproduce there without destroying the protoplasmic unity of the host, and, finally, the extent and rapidity of the regressive changes in which the parasitic microbe might indulge in view of the great opportunity offered. That visible microbes are capable of penetrating cells and reproducing in them is seen in the case of the rickettsial organisms that cause typhus and spotted fever. The faculty of microbes to acquire the character of growing to abundance within the protoplasm of a cell without causing excessive damage to the cell is also shown by the *Rickettsia*. These visible microbes occur in enormous numbers within the protoplasm without destroying the contours and the finer structure of the host cell. There is apparently some reason why so few visible microbes are found living an intracellular existence. In general, disease-producing microbes live between cells, and viruses live within cells. We may suggest that when a visible bacterium has developed the ability to grow in protoplasm, its opportunity for retrogradation becomes so great that simplification and decrease in size take place with relative rapidity as an evolutionary process and the microbe is converted into the incomplete form called a virus.

*Further observations on the microbic origin of viruses.*—Like that of evolution, the theory of the transformation of microbes into viruses cannot be proved experimentally or by any one set of facts. Acceptance of this concept must come from a general agreement that all of the related facts observed in nature confirm the view. The concept of a microbic origin of viruses is supported not only by the adequacy of the explanation it offers but also by the absence of any other reasonable explanation. It offers an explanation of the nature as well as the origin of viruses.

Processes of simplification or degeneration under the stimulus of parasitism are observed generally in forms of life larger than the microscopic. That unicellular forms of life have the plasticity required for adaptive change is easily deduced. The

## ROBERT G. GREEN

complex organelles developed in protozoa show the high degree to which protoplasm itself may be specialized. The adaptation of the temperatures of pathogenic bacteria to the temperature of their host and their dependence upon a host for certain essential elements of metabolism demonstrate that the protoplasm of the microscopic forms has the plasticity essential to simplification. It becomes entirely probable that the retrograde anatomical changes observed in larger forms of life occur in microscopic forms of life also. The theory of a microbic origin of viruses leads us only one step further: that is, to recognition that the processes of regression may continue until they bring about the utmost simplification of a microbe and reduction of its size to a point beyond microscopic visibility.

In the light of the concept of a microbic origin of viruses many characters of viruses seemingly inexplicable are adequately explained or are subject to logical consideration. As previously noted, few visible bacteria pathogenic for man and animals are found within cells. Ordinary pathogenic microbes typically grow between cells rather than within them. Rickettsia, the group of visible microbes described above as growing characteristically within cells, are like bacteria in that they are visible under a microscope and have the general appearance of microbes; they are like viruses, however, in that they are extremely minute, being just within the limit of visibility. Their staining qualities are in many respects like those of the inclusion bodies associated with virus infections. Most important, they are like viruses in that they will not grow on an artificial medium, even one to which has been added the tissue fluid or juice that allows the pathogenic bacteria to multiply. The Rickettsia evidently grow only in actually living cells, as do viruses. They seem to lie halfway between visible microbes and filterable viruses, if one takes into consideration all of their major characters. The Rickettsia have an additional, special character in that they are all transmitted by insects or arthropods. They are highly adapted to living within these hosts as well as in the animals in which they produce disease. Indeed it is not unlikely

## THE BIOLOGY OF ANIMAL VIRUSES

that they are adapted to living primarily in arthropod and insect hosts and that their growth within an animal provides a means of transfer from one blood-feeding insect host to another. There is probably considerable difference between the metabolism of the cells forming ticks and that of the cells forming mammals. Consequently, any retrograde or degenerative processes that take place in the *Rickettsia* would have to be duplicated in two widely diverse hosts and would follow a slower evolutionary process than do adaptations to a single form. It might well be that a point would be reached in certain combinations of animal and arthropod hosts at which an adaptation to an animal could not be matched by a similar adaptation to an arthropod, and a fixation might result that would prevent further progress. According to this view the *Rickettsia* are forms intermediate between microbes and viruses and persist as a singular group because they are anchored by divergent complexities in two unrelated hosts.

*A biological concept of virus infections.*—Because of their small magnitudes, viruses cannot be seen. All we see of a virus is the effects which it produces. Like the Cheshire cat in *Alice in Wonderland*, which faded away and left its grin, certain pathogenic microbes seem to have faded from view and left their diseases to puzzle us. If, as I believe, viruses are incomplete microbes that originate from protozoa and bacteria, they are biological units; and their conduct in an animal body may be interpreted as that of a species in maintaining itself.

The activity of the causative agent is much more important in the study of a virus disease than in the study of a microbic disease. We can see the inciting organism of a microbic disease and may therefore overlook details of its life history. In reality we should not overlook them, but because we are able to see the organism we are fairly successful in combating it without a thorough knowledge of its life history. Yet, since procedures for the control of a virus disease must be based entirely on a knowledge of the effects which the virus produces, we should thoroughly understand its life history.

## ROBERT G. GREEN

As a biological unit, a virus is primarily a highly specialized kind of plant or animal that attempts to maintain itself in accordance with the same fundamental laws and principles by which other species of life maintain themselves. The host infected by them becomes the environment in which populations of the virus increase or decrease in proportion to the opportunities afforded. A virus that gains entrance to an animal body is accorded opportunities comparable to those enjoyed by a rabbit entering a virgin stand of timber after a decimation of rabbits. The rabbit may live and feed in the manner to which it is accustomed or adapted. Depending upon its species, it may prefer to live entirely on the ground or to make its home in burrows under the ground. It may select leaves and shoots for food or it may select buds. Some plants it will consistently eat, some it will shun. An understanding of the activities of viruses is derived from a similar consideration of them as biological species. When a virus enters an animal body it finds itself in a vast and complex environment, a world in which it may produce a cyclic population. Within certain limits the virus leads the orderly life that is more or less prescribed by nature for the perpetuation of its species. It lives and grows in certain anatomical locations and in certain types of cell; other locations and types of cell it shuns.

The virus of dog distemper seeks primarily as its habitat epithelial cells in a surface position. In an infection of dog distemper, cellular changes caused by the virus are found in the epithelial cells covering the eyeball, in the membranes of the nose and throat, and in the lining of the bronchial tree. The virus may likewise be identified in the epithelial cells that make up the lining of the bile excretory system and in the surface epithelial cells of the urogenital tract, especially those of the pelvis of the kidney, the ureters, and the bladder. Other types of cell the virus of dog distemper may enter as second choice, such as the nerve cell, which is of epithelial origin. The virus of dog distemper rejects entirely, however, many groups of cells, such as the hepatic cells that constitute the bulk of the liver.

## THE BIOLOGY OF ANIMAL VIRUSES

The fox encephalitis virus invades cells that make up the lining of the blood-vascular system, especially those surrounded by nervous tissue. This virus secondarily accepts the hepatic cells of the liver as a habitat.

Adaptations of viruses to types of cell supersede adaptations to anatomical locations in the individual host and, within certain limits, even supersede adaptations to the species of animal invaded. Although the distemper virus from the surface epithelium of a dog does not grow in the liver cells of a dog, it does find a suitable environment in the surface epithelial cells of such diverse species as the fox, the raccoon, the mink, and the ferret. It is becoming obvious that the requisite habitat of filterable viruses is the cell and that the adaptation of viruses to cells is more fundamental to a virus-host relationship than is growth of the virus in particular organs or in certain species of animal. Only a full appreciation of these relationships can stimulate the extensive research needed to work out in detail the distribution of viruses in the various cellular systems in all susceptible hosts.

Not all of the activities of viruses are directed by compulsion of adaptation for species preservation. In the case of certain viruses some of the most profound effects produced in the environment may have no bearing at all on the welfare of the species but may be purely fortuitous. It has not yet been shown, for example, that the penetration of the virus of infantile paralysis into the cells of the spinal cord is essential to the maintenance of this virus in nature. It may be an accidental circumstance that follows growth of the virus in the upper respiratory tract or in some other superficial location. Moreover, we cannot even say at this time that the human infection with this disease is important in maintaining the causative virus in nature. There are certain features of epidemics of poliomyelitis in human beings which indicate that the infection in them is wholly accidental and incidental to growth and maintenance of the virus in some entirely different species of life.

Not until we are acquainted with the complete life history of viruses will we understand the diseases produced by them. In

order to control virus diseases we must have an exact knowledge of those activities of the virus that are essential to its maintenance as a species and of those that are accidental. The rigid prerequisites for the survival of a virus are twofold: first, certain habits must be of such frequency as to insure transfer from one host to another, thereby perpetuating the species; second, fortuitous activities that hinder the transfer to a new host must not be so frequent as to result in failure of host-to-host transfer and thus destroy the continuity of reproduction.

*Life histories of filterable viruses.*—The life history of a filterable virus, then, must reveal a full knowledge of the activities of the virus that have to do with its survival and modification in nature, and must clearly define as such fortuitous, secondary habits of existence, which may sometimes be those that give rise to the most distressing features of a clinical disease.

It may be doubted that the term *life history* should be accorded general usage in the developing science of filterable viruses. However, the concept implied is essential to a full appreciation of the nature of viruses and the diseases caused by them. Throughout the field of pathogenic microbiology, certain terms such as *portal of entry* and *portal of exit* have long been used in descriptions of the infectious-disease process. These terms are critical in a concept of life history of viruses. They are not commonly used, however, in studied reference to survival of the organism and its environmental relationships.

We do not yet have enough information on any virus disease to enable us to construct an adequate life history of the virus, if by adequate we mean sufficient for both scientific understanding and control of the infection. The virus disease of human beings for which the best preventive measures are available is smallpox. Periodic, if somewhat irregular, inoculations of human beings with the cowpox virus usually hold this disease in check, below epidemic proportions, in the United States, and control it to a greater or less degree in other parts of the world. When we consider that the jennerian vaccination was discovered a hundred years before a filterable virus was first demon-

## THE BIOLOGY OF ANIMAL VIRUSES

strated, we perceive how little relation there may be between scientific understanding of a disease and successful methods of prophylaxis. It is only recently that investigations have been conducted on smallpox that are comprehensive enough to yield eventually the scientific knowledge needed to construct a life history of the virus of this infection. The development of smallpox vaccination without a thorough scientific study of the infection stands alone in the annals of medicine. Intensive researches conducted for years on other virus infections have failed to yield the information required for prevention or cure of the infections. The future of a science of virus diseases demands systematic information concerning all virus diseases even though their control may be accomplished long before a life history of the viruses is obtained.

In our laboratories we have attempted over a period of years to discover and explain facts of the life history of two viruses that produce disease in canines. One of them, the virus of fox encephalitis, we have found to produce recognizable infections in canines only; the other, the virus of canine distemper, we have found to be infectious not only for dogs but also for ferrets, foxes, minks, and raccoons. The latter virus, experimental work indicates, may likewise be infectious for many additional species related to those mentioned. Only a brief summary of our most significant results may be given here.

The virus of fox encephalitis produces intranuclear inclusions, which, in our judgment, identify the cells invaded by the virus. By measuring the cubic content of microscopic sections and determining the actual number of inclusions present, it has been possible to map quantitatively the distributions of virus in the tissues of foxes during infections.

In quantitative studies on ten naturally infected foxes and ten foxes inoculated experimentally, 22 inclusion bodies to the cubic millimeter were seen in the brain and organs of the naturally infected animals and 24 inclusions to the cubic millimeter in the brain and organs of the experimentally infected animals. In the natural infection of fox encephalitis 91 per cent of the inclu-

## ROBERT G. GREEN

sions were situated in endothelial cells, 7 per cent in the pia-arachnoidal cells of the membranes covering the central nervous system, and 2 per cent in liver cells. In the experimentally infected foxes only 59 per cent of the inclusions were in the endothelial cells of blood vessels, 30 per cent in the pia-arachnoidal cells, and 3 per cent in the ependymal cells lining the cavities of the brain. In this group 8 per cent of the inclusions were in the hepatic cells of the liver. These figures tell us that the virus of fox encephalitis has as its seat of primary growth the endothelial cells that comprise the blood-vascular system. An increased growth of the virus in the pia-arachnoidal cells is produced if the virus is artificially injected into them in large quantities. The figures reveal also that the virus does not ordinarily penetrate the ependymal cells that make up the lining of the cavities of the central nervous system. These cells are susceptible to virus invasion, however, and become affected if the virus is placed directly in contact with them.

Further analysis suggests that in the natural infection the activity of the fox encephalitis virus is equally pronounced in the cerebral hemispheres and in the brain stem, producing an average of 25 inclusion bodies to the cubic millimeter in each. The growth of the virus appeared greatest in the basal ganglion and least in the pons. Inclusion bodies were absent almost altogether from endothelial cells of the spinal cord. A comparison of the naturally infected foxes and the experimentally infected foxes shows that fewer inclusions were present in the brain of the experimentally infected animals. However, in the liver cells there were more than two and one-half times as many inclusion bodies in the experimentally infected animals as in those naturally infected. From these data it is evident that the infection is naturally a disease of the brain but that it becomes less an infection of the brain and more a disease of the liver when it is transmitted experimentally through a series of animals. From such studies as we have conducted on the disease in dogs it appears that in these animals the infection is even more a disease of the liver. Whether this is true cannot be accurately deter-



## THE BIOLOGY OF ANIMAL VIRUSES

mined until additional quantitative studies are carried out comparing the infection in dogs and foxes.

In the case of fox encephalitis we see a virus that accomplishes its reproduction mainly in a very specific group of cells, endothelial cells, distributed throughout the animal body, and particularly in endothelial cells located in the brain. The second type of cell in which this virus reproduces by preference appears from our studies thus far to be the hepatic cells of the liver. We have observed that this virus affects a limited number of cells in other systems, especially in the reticulo-endothelial system. We do not know whether or how the growth of this virus in the endothelial cells of the blood vessels aids the virus in survival; its growth in these cells may be an irrelevant occurrence. Experimentally the virus has been isolated from the secretions of the upper respiratory tract both in the natural infection and in the experimental infection. It would seem that the disease is transmitted through the upper respiratory tract. We have found large numbers of inclusions forming the base of an ulcer in the tonsil. In this case the cells containing the inclusions are to be classed as reticulo-endothelial. Only a single inclusion of fox encephalitis has been found in what appeared to be a surface epithelial cell. We have as yet learned little that allows a true understanding of the nature of this infection even though procedures have been perfected that control the disease on fox farms. We do not understand fox encephalitis as a disease because we have not yet become sufficiently well acquainted with the life history of its virus.

Considerably more information is available relating to the life history of the canine distemper virus. Since the work of Laidlaw and Dunkin, in which they showed that the virus is common to dogs and ferrets, we have identified canine distemper in foxes, raccoons, and minks. The presence of this virus is made known by two types of inclusion. The more common and characteristic are the cytoplasmic inclusions, which resemble closely the Negri bodies of the rabies virus. Although both the fox encephalitis virus and the distemper virus produce intra-

## ROBERT G. GREEN

nuclear inclusions, they do not produce them in the same type of cell. The intranuclear inclusions of canine distemper are found in epithelial cells. In all susceptible species of animal the virus of canine distemper avoids the endothelial and the liver cells, which we have seen are utilized more than other cells by the fox encephalitis virus.

The distemper virus, as even qualitative study will affirm, reproduces principally in the epithelial cells of surface membrane, although inclusion bodies caused by this virus are found in some cells of the deeper tissues, such as reticulo-endothelial cells and chromatin cells of the adrenal. In tracing the life history of this virus it seems obvious that the marked involvement of surface epithelium indicates the effective reproduction of the virus in maintaining itself. The origin of the virus in the upper respiratory secretions, in the urine, and in the feces, if it there occurs, is made plain by the destruction of epithelial cells lining the associated organs. A question may be raised whether the large amount of distemper virus noted in the spleen is actually produced there or whether it is strained out of the circulatory system and concentrates in the spleen. There may be a function related to reproduction of the virus in the deeper cells of the body, such as the reticulo-endothelial cells of the spleen. Virus produced in the deeper locations may enter the blood and assist in a complete seeding of the surface epithelium of the body.

The growth of the distemper virus in surface epithelium seems to represent the reproduction of the virus essential for its maintenance in nature. As multiplication of the virus occurs in superficial cells, both entrance to and exit from a host are accomplished by direct contact. It is evident, then, that in general the facts established concerning distemper as an infection in the several animals studied actually contribute an authentic chapter on the reproduction of the distemper virus to its life history.

Our more recent investigations on the distemper virus furnish a wholly new section in its life history. We have found that this virus is modified in various ways by the different environments in unrelated species of animal and that its possibilities for main-

## THE BIOLOGY OF ANIMAL VIRUSES

tenance in nature are greatly affected thereby. The distemper virus has the ability to grow in animal species throughout a rather broad zone of zoological relationships. Yet the ability of the virus to produce infections successfully in any one species and to pass from one species to another is greatly altered by its successive residences in individuals of one species. A distemper virus that has been passed through ferrets fifty or more times serially has a greatly increased virulence for the ferret, but is able to produce only a mild, symptomless infection in foxes. Such a virus is somewhat analogous to the fixed rabies virus of Pasteur, and probably more comparable to the vaccinia virus when it is considered as a smallpox virus which has been passed serially through calves and become modified in its relation to human infection. I have termed the modified distemper virus a *distemperoid virus*. The modified distemper virus seems to possess a changed virulence for the various groups of susceptible animals which is interrelated systematically with their zoological relationships.

The ability of an ultramicrobe to invade a number of animal species, and especially unrelated species, is a distinct advantage and must represent a specific adaptation. From work done on distemper thus far it is known that this disease is found in members of the family Mustelidae, which includes the ferret, mink, and related animals; in the family Canidae, which includes the dog, wolf, and red fox; and in some intermediate species such as the raccoon. Both experimental and field studies show that the distemper virus develops characters in its passage from one individual to another of the same species that determine its ability to maintain itself in that species or to spread from one species to another. A true understanding of distemper, either as a disease or as the activity of a species of microbe maintaining itself in nature, must be built in part on a knowledge of this phase of the life history of the virus.

The virulence of the distemper virus as related to animal species is depicted in Figure 3. If a distemper virus that is highly pathogenic for both ferrets and foxes is passed serially through

## ROBERT G. GREEN

ferrets, it is transformed into a modified virus which, while highly pathogenic for ferrets, is of very low virulence for foxes. If a distemper virus that is highly pathogenic for ferrets and foxes is passed serially through foxes, it is transformed into a distemperoid virus of very low virulence for the ferret. The

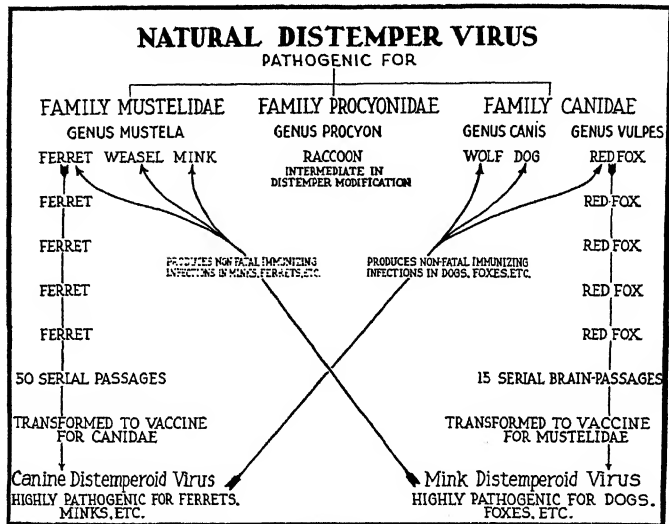


Figure 3. Modifications in virulence of the distemper virus.

modification to increased virulence is effective for closely related species. Canine distemper virus that has been passed serially through ferrets is highly virulent for ferrets and also very virulent for minks, but less so for raccoons. Similarly the virus that has been passed serially through foxes is highly virulent for dogs, less so for raccoons, and almost nonpathogenic for minks and ferrets. Modifications in virulence of the distemper virus that occur in the family Mustelidae and in the family Canidae might be described as mirror images of each other. The raccoon, being intermediate between the ferret and the fox, is moderately

## THE BIOLOGY OF ANIMAL VIRUSES

susceptible to invasion by viruses modified either by passage through ferrets or by passage through foxes.

*Changes in virulence of the distemper virus in nature.*—These modifications in the distemper virus, which have been experimentally demonstrated, appear to be a reflection of the fluctuations in virulence of the distemper virus that occur in nature. That distemper is a disease of wild animals in their natural environment we have shown by the isolation of the distemper virus from an epizootic in wild foxes and from an epizootic among raccoons in the wild state.

The usual type of distemper virus in nature seems to have the ability to produce infections in such diverse species as the fox and the mink. Strains of this type isolated from foxes and dogs produce typical distemper when injected into ferrets or minks. A distemper virus capable of spreading from one species to another, producing severe infections, might be termed a normal distemper virus. This type is to be considered the original type of virus developed among animals free in the wild, where the disease spreads readily from one species to another.

The congregating of many animals of the same species, as has been accomplished by man, leads to definite changes in the distemper virus. Whereas, in the wild, distemper might easily spread from a weasel to a fox, from the fox to a raccoon, and on to other species, distemper producing an epizootic on a fox farm spreads only from fox to fox. We have shown experimentally that as few as five passages through a single species of susceptible animal produce a discernible change in the pathogenic properties of distemper virus. Distemper that has spread through a fox farm over a period of years has become greatly modified.

It has been most puzzling to see distemper cause a high mortality among foxes and not spread at all to minks on a ranch where both animals were present. The explanation is now very plain. In cases of this kind the particular strain of virus that spread to the ranch was already highly adapted to foxes and had probably come from a long series of infections that passed through dogs and foxes only, probably chiefly through foxes.

## ROBERT G. GREEN

The virus had thus become so modified that it was incapable of producing a severe infection in minks. As it was carried to a ranch on which both foxes and minks were raised, it spread rapidly through the foxes as a virulent infection, and very slowly, if at all, through the minks as a mild infection.

The opposite situation, in which disease appears among minks and does not spread to foxes, has a similar explanation. The distemper virus in this case is a strain that has previously passed through minks, ferrets, weasels, or similar animals, and become so modified that it produces only a mild infection or no infection in foxes although it produces a virulent infection in the minks.

Information available from field observation and experimental work makes it appear certain that the pathogenic properties of the distemper virus found in nature are dependent upon previous experiences of that virus. In modifications which make it possible for the distemper virus to affect animals ranging from the ferret and related species to the red fox and related species, intermediate species, such as the raccoon, act as a vehicle. For example, if a virus became so highly adapted to members of the family Mustelidae that it could not maintain itself in the family Canidae, passage of the virus through raccoons or other members of intermediate families might so transform the virus that it could obtain a foothold in some members of the family Canidae and undergo adaptations that would again stabilize the virus as an invasive agent for members of that family. Such transformations of virulence bear directly on the movement of the virus from host to host and on the degree of reproduction in each host. They are essentially concerned with the survival of species and must be of great importance in the life history of a virus.

# THE MODE OF ACTION OF SULFANILAMIDE AND ITS DERIVATIVES

BY

Perrin H. Long

*Associate Professor of Medicine,  
Johns Hopkins University*

IT HAS seemed strange to some (1) that although more than six years have elapsed since Foerster (2) first reported upon the clinical use of Prontosil in the treatment of a staphylococcal infection, there is still doubt in the minds of many observers concerning the mode of action of sulfanilamide and its derivatives. However, it is to be remembered that, despite more than twenty-five years of intensive investigation, pertinent facts regarding the action of arsphenamine upon the spirochaete are just coming to light, and that knowledge of how and by what mechanisms chemotherapeutic agents in general are effective is very scant. In view of our lack of information on the fundamental processes involved when an effective chemical agent injures or kills a susceptible microorganism, it is scarcely surprising that we cannot explain in definite terms the way in which sulfanilamide and its derivatives behave in the control of infectious diseases.

In Domagk's (3) original report upon the chemotherapeutic effects of Prontosil in experimental hemolytic streptococcal infections in mice, the fact that the drug was effective *nur im lebenden Organismus* was stressed, as was also its lack of *bacteriostatic* or *bactericidal* effect in vitro. This observation has been repeatedly confirmed (4-7) in respect to both Prontosil and its companion azo-dye, Neoprontosil.

This divergence between the in vitro and in vivo activity of Prontosil puzzled the earlier observers and led to the formation of certain hypotheses regarding the action of the drug. Levaditi and Vaisman (8) first considered the possibility that the drug

activated host defense mechanisms, next that it prevented the formation of capsules, thus rendering the virulent streptococci susceptible to phagocytosis (8), and finally, late in 1935 (9), that the drug probably acted upon susceptible microorganisms in a way which lessened their defenses against those of the host. Domagk (3, 10-12), on the contrary, held that Prontosil per se or a conversion product of this drug acted directly upon susceptible bacteria in vivo. In his latest report this observer (13) stated that "the first phase of action consisted always of an attack directly upon the germs, but the microbes were rarely completely destroyed, being more often sufficiently modified to be destroyed by means of the defense mechanism of the host." Thus, according to Domagk, the streptococci in Prontosil-treated animals behaved like avirulent streptococci.

Little real progress in this problem was made until late in 1935, when the Tréfouëls, Nitti, and Bovet (14) announced that it was their belief that Prontosil broke down in the animal body to triaminobenzene and para-aminobenzene sulfonamide (sulfanilamide) and that they had found the latter compound to be an active chemotherapeutic agent in the control of experimental streptococcal infections in mice. This, obviously, was an observation second in importance only to that of Domagk, because it showed that the azo linkage was unnecessary for therapeutic activity and provided the needed stepping stone for further investigations of the mode of action of these compounds.

However, as has been pointed out, "to show that sulfanilamide is an effective chemotherapeutic agent is a far call from proving that it is the active component of Prontosil" (15). Nevertheless, evidence was soon produced that this might be the case when Fourneau, the Tréfouëls, Nitti, and Bovet (16) reported that the addition of sulfanilamide to cultures of *Aspergillus niger* delayed the growth of these organisms in Raulin's medium, thus demonstrating for the first time the in vitro bacteriostatic activity of sulfanilamide. This observation is the keystone upon which rests the major structure of subsequent work dealing with the mode of action of sulfanilamide and its



## SULFANILAMIDE AND ITS DERIVATIVES

derivatives. It has been confirmed by numerous observers working with a wide variety of microorganisms.

Another link in the chain of evidence that sulfanilamide was the active molecule in Prontosil was the demonstration by Colebrook and his associates (17) and by ourselves (18) that the reduction of Neoprontosil *in vivo* or *in vitro* results in the production of a bacteriostatic substance and, as was shown by Fuller (7), that the effective material liberated both *in vitro* and *in vivo* from the Neoprontosil molecule in the course of reduction is sulfanilamide.

TABLE 1. — THERAPEUTIC EFFECT OF BRIEF TREATMENT WITH SULFANILAMIDE AND ITS WEIGHT OR MOLAR EQUIVALENT OF NEOPRONTOSIL UPON EXPERIMENTAL HEMOLYTIC STREPTOCOCCAL INFECTIONS IN MICE

Compound	Route of Therapy	No. of Mice	No. of Deaths from $\beta$ Hemolytic Streptococci (by days)														Survivals
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	No. %
Sulfanilamide, 6 mg.....	Parenteral	150	..	18	17	27	30	6	3	1	2	..	1	1	..	..	44 29
Neoprontosil, m.eq., 20.5 mg.....	Parenteral	100	7	6	6	19	14	9	2	1	2	..	..	..	1	..	33 33
Neoprontosil, w.eq., 6 mg.....	Parenteral	100	29	22	12	15	6	1	..	..	1	..	..	..	..	..	14 14
Sulfanilamide, 10 mg.....	Peroral	50	..	12	6	11	12	4	..	..	..	..	..	..	..	..	5 10
Neoprontosil, 10 mg.....	Peroral	50	9	27	10	1	1	1	..	..	..	..	..	1	..	..	0 0
Neoprontosil, 36.2 mg.....	Peroral	50	2	2	3	16	14	7	..	..	..	..	..	..	..	..	6 12

Further data bearing upon the assertion that an azo-dye such as Neoprontosil is active only in proportion to its potential liberation of sulfanilamide is offered in Table 1, which shows the therapeutic results obtained when mice infected with  $500 \pm$  lethal doses of the C203 strain of hemolytic streptococci were treated with equal weights and molar equivalents of sulfanilamide and Neoprontosil by the parenteral and peroral routes. The treatment was instituted immediately after infection and was repeated once a day for three days. It is to be noted that when the drugs were administered by the parenteral route, the

therapeutic effect of the molar equivalent dose of Neoprontosil was essentially that of sulfanilamide, while the results obtained from the weight equivalent dose of Neoprontosil were much inferior. The same was true when these drugs were administered by the peroral route.

At this point it is of importance to point out that there have been numerous objectors to the theory that Prontosil and Neoprontosil are active only by virtue of their reduction to sulfanilamide. Domagk (10, 11) opposed this belief, first, because he was unable to find any great difference in the *in vitro* bacteriostatic activity of Prontosil and sulfanilamide (an observation at variance with the experience of practically all other observers and due to marked technical differences in the conduct of his experiments); secondly, because he considered that the presence of sulfanilamide in the urine of patients receiving Prontosil was no indication that it was the active fraction of the dye compound; and finally, because the liberation of sulfanilamide during the course of Prontosil therapy was not proof that it was the active fraction, inasmuch as there were a number of therapeutically active compounds which could not possibly liberate sulfanilamide. This last argument is extremely weak since no one has claimed that sulfanilamide is the only conceivable active compound.

While we do not intend to dwell at length upon the relation of chemical structure to therapeutic activity, a few observations are worth mentioning for the light they throw on this subject.

It seems quite likely, as we have just pointed out, that the activity of the azo compounds is dependent upon the destruction of the  $N=N$  linkage in the body and the subsequent formation of sulfanilamide. The amino group in benzene sulfonamide compounds must be para to the sulfonamide group. Meta- and ortho-amino compounds are inactive. Substitutions in the amino group of sulfanilamide generally lower the effectiveness of the compound; and, as far as our observations go, the activity of any sulfanilamide derivative containing the  $NH-R$  group

## SULFANILAMIDE AND ITS DERIVATIVES

is dependent upon the splitting-off of the R radical, and the subsequent reversion of the compound to sulfanilamide. (The possible exception to the rule would be the hydroxylamine derivative.) It has also been demonstrated that the para amino group is not necessary for activity, as para-nitrobenzene sulfonamide is an effective compound. (The stability of this derivative in vivo is still a matter of discussion.)

Substitutions in the sulfonamide group ( $\text{SO}_2\text{NH-R}$ ) may or may not result in a compound showing a lowered degree of activity. For example, sulfapyridine is a highly effective compound, while N(p-aminobenzene sulfonyl)-1, 4-oxazine is practically without activity. As far as is known, most of the compounds resulting from substitutions made in the sulfonamide group are stable and, if active, derive their effect from the whole molecule. In other words, the  $\text{SO}_2\text{NH-R}$  linkage is firm. It has also been demonstrated that the  $\text{SO}_2\text{NH}_2$  group is not the only active radical and that the S, SO, and  $\text{SO}_2$  groups possess varying degrees of activity, provided they are placed para to the amino group. Recently Mayer and Oechsli (19) have shown that sulfur-containing radicals are unnecessary for activity, by demonstrating the chemotherapeutic effectiveness of para-nitrobenzoic acid in the treatment of experimental streptococcal infections.

Therefore we can assume from what is known that the molecules of sulfanilamide and sulfapyridine, while possibly undergoing oxidative changes in the body (at least according to certain theories regarding their mode of action), do not suffer radical alterations in their chemical structure during their period of activity in the body of the infected host. Hence it is highly likely that the way in which sulfanilamide acts upon susceptible microorganisms is similar to the action of sulfapyridine, the variations being only in the intensity of action.

In discussing the mode of action of sulfanilamide and sulfapyridine, it should be borne in mind that while their activity appears to be directed solely against the invading microorganism, the recovery of the infected subject seems to entail two

## PERRIN H. LONG

factors — the drug factor and the host factor. That is to say, while the drug can bring the infection under control, it requires the cooperation of the host's defense mechanism to dispose of the infectious agent. This statement is in agreement with the conception of those investigators who have attacked the problem by studying the *in vitro* and *in vivo* effects of these drugs.

Under the heading of the drug factor comes the activity of these compounds in inhibiting the growth of, or possibly actually killing, susceptible bacteria, and their ability to neutralize the harmful effects of the toxic products of certain microorganisms. The host factor represents the response of the body's defense mechanisms, such as antibody production and mobilization of phagocytic cells, to the infection produced by microorganisms which have been altered as a direct result of the drug factor.

At the present time it is possible to select from the welter of conflicting opinions three main hypotheses concerning the manner in which sulfanilamide and sulfapyridine exert their bacteriostatic effect upon susceptible bacteria. The first of these is based upon data which are believed to indicate that oxidative changes, in which molecular oxygen plays a role, are responsible for the conversion of sulfanilamide or sulfapyridine into "truly" active compounds.

Early in 1937 Mayer (20) concluded that as sulfanilamide appeared to be less active *in vitro* than *in vivo*, one must postulate the formation of a more active compound in the body. This he claimed must be an oxidation product (the process being brought about by an oxidant in the body). In support of his hypothesis, Mayer cited the frequent appearance of methemoglobinemia in patients receiving sulfanilamide, an occurrence which, according to him, implied the presence of an agent capable of oxidizing hemoglobin. Sulfanilamide would be unable to do this, but its first oxidation product, para-hydroxylaminobenzene sulfonamide, could bring about this change. He next synthesized para-hydroxylaminobenzene sulfonamide and reported that its bacteriostatic activity was about one hundred

## SULFANILAMIDE AND ITS DERIVATIVES

times that of sulfanilamide in vitro. In addition, Mayer stated that other oxidation products of sulfanilamide, such as the azoxy, nitro, and nitroso derivatives, were also highly active compounds. Little, however, was done immediately to prove or disprove Mayer's hypothesis because of the difficulties surrounding the preparation of the hydroxyl derivative of sulfanilamide.

About a year after Mayer's report was published, Ottenberg and Fox (21) noted that irradiated dilute solutions of sulfanilamide turned blue. Later they (22) reported that this phenomenon did not occur in the absence of oxygen and that the blue substance could be reduced to a colorless product, thus indicating the existence of an oxidation-reduction system. However, they were unable to demonstrate that the blue substance possessed bacteriostatic activity or that the reduction product was sulfanilamide.

Recently Fox, German, and Janeway (23) studied the effects of the addition of sulfanilamide upon electrode potentials in sterile broth (which were unaffected) and in cultures of hemolytic streptococci. They found that whereas the electrode potential in such cultures fell rapidly during normal growth, it remained elevated during the time that sulfanilamide was exerting its bacteriostatic effect. In the presence of cysteine, or when air was excluded by sealing the cultures with vaseline, the potential was lowered despite the addition of sulfanilamide, and the bacteriostatic activity of the drug was definitely decreased.

Warren, Street, and Stokinger (24), unlike Fox and his associates, did observe an elevation of the potential of sterile broth to which sulfanilamide had been added, but they doubted that the drug poisoned the system at a "critical"  $E_h$ . They too found that the electrode potentials of cultures of streptococci fell less rapidly in the presence of sulfanilamide. They noted that when the drug was added to 21-hour-old cultures, a rapid rise in potential occurred, and that under anaerobic conditions there was no difference between the potentials of the control and sulfanilamide cultures. The same was true in the presence of cysteine,

## PERRIN H. LONG

and in such cultures the bacteriostatic activity of sulfanilamide was reduced to a minimum. These investigators did not reach any definite conclusions as a result of their observations, but suggested that it was possible that sulfanilamide inactivated enzyme systems by attacking sulfhydryl or similar groups which are "normally responsible for the attainment of highly negative potentials."

Quite recently Shaffer (1) has advanced the hypothesis that sulfanilamide is oxidized in the presence of hydrogen peroxide, plus essential catalysts, to an active compound. This product is so strong an oxidant that it destroys catalase, thereby permitting more hydrogen peroxide to accumulate. As a result more and more of the oxidation product of sulfanilamide is formed, until eventually a concentration is reached sufficient to attack all reactive reducing systems of the cell, and bacteriostasis or actual killing of the microorganisms ensues.

The second hypothesis dealing with the mode of action of sulfanilamide was advanced a year ago by Locke, Main, Mellon, and Shinn (25-31). These observers noted that dilute solutions of sulfanilamide which had been irradiated with ultraviolet light possessed the property of inactivating catalase. It has been long known that if catalase is inactivated in aerobic cultures of pneumococci, peroxide accumulates rapidly and may reach a concentration that is bactericidal for the pneumococcal cells. Hence they reasoned that if sulfanilamide was able to inactivate catalase, peroxide would accumulate, and this factor might account for the bacteriostatic or bactericidal action of sulfanilamide.

They described the mechanism as follows: "The growing bacterial cell has the power to convert sulfanilamide, presumably through mild oxidation, into a derivative which is a highly active anticatalase. This reaction produces an accumulation of anticatalase in the immediate vicinity of the cell. The streptococcus and pneumococcus, being active producers of hydrogen peroxide, are able to grow only so long as the peroxide concentration can be kept below a critical level by outward diffusion

## SULFANILAMIDE AND ITS DERIVATIVES

or destruction. . . . in the presence of anticalase, inactivation of catalase takes place in the zone immediately adjacent to the cell with resultant accumulation of hydrogen peroxide to toxic levels."

In support of this view they brought forth evidence to show that peroxide accumulated to higher concentrations in pneumococcal cultures containing sulfanilamide than in cultures without sulfanilamide, and that if there was a "reduction of the percentage of oxygen in the superambient air of broth cultures of the type I pneumococcus," the bacteriostatic effects of sulfanilamide were either greatly reduced or altogether prevented. However, they also found that when the oxygen concentration was reduced to below 0.04 per cent (not even approaching anaerobiosis!) sulfanilamide was again a strongly bacteriostatic agent in cultures of pneumococci. They further reported that they had observed chain formation in sulfanilamide-containing cultures of pneumococci and suggested that "the nature of this change is reminiscent of the morphology assumed by the avirulent R. culture phase of the pneumococcus."

Before discussing the third explanation offered for the mode of action of sulfanilamide, we should like to point out that the two hypotheses just discussed are based upon a supposed oxidation of sulfanilamide to an active derivative and that an important point in support of this belief is the claim that sulfanilamide and sulfapyridine are inactive under anaerobic conditions. Hence if it could be shown that sulfanilamide and sulfapyridine are effective bacteriostatic agents under conditions of strict anaerobiosis, this would constitute a definite objection to both the oxidation and anticalase theories, since both are based upon the assumption that molecular oxygen is necessary for the reaction.

Early in 1937 we (18) demonstrated that the reduction of Neoprontosil could be brought about by the addition of an excess of cysteine hydrochloride and that the resulting sulfanilamide was bacteriostatic against the C203 strain of hemolytic streptococci even though marked reducing conditions

TABLE 2. — EFFECT OF ANAEROBOSIS ON THE BACTERIOSTATIC ACTION OF SULFANILAMIDE AND SULFAPYRIDINE, SHOWING THE DEGREE OF BACTERIOSTASIS \* AFTER 16-22 HOURS AT 37° C

Medium †	Type of Anaerobiosis ‡											
	Experiment 1			Experiment 2			Experiment 3			Experiment 4		
	Jar	Seal	None	Jar	Seal	None	Jar	Seal	None	Jar	Seal	None
C203 STRAIN OF GROUP A HEMOLYTIC STREPTOCOCCI												
Cysteine +	238	844	....	52	7	...	10	3.6	2.7	1,333	>100	300
P.A.S. ....												
Cysteine +	006	775	....	134	2,885	...	80	33	18	1,250	>1,000	2,183
S.P. ....	9	32	63	3.8	2.5	484	3	10	6.3	2	50	19
P.A.S. ....												
S.P. ....	226	2,500	8,533	10	10	187	11	112	386	88	200	231
TYPE I PNEUMOCOCCI												
Cysteine +	...	2.5	....	0	4	...	1,272	755	>5	1,379	524	9
P.A.S. ....												
Cysteine												
+ S.P. ....	...	...	...	...	...	...	...	...	...	...	...	...
P.A.S. ....	...	1,315	80,645	2	7,570	8.6	1,085	1	145,000	345	750	483,333
S.P. ....	...	...	...	...	...	...	...	...	...	...	...	...
Averages												
	Jar	Seal	None	Jar	Seal	None	Jar	Seal	None	Jar	Seal	None
	278	84	77	641	766	561	27	75	109	436	583	1,202

\* Degree of bacteriostasis obtained by dividing growth in control by growth in test = growth in control

† Medium = beef infusion plus 2 per cent Neopeptone and 0.075 per cent dextrose plus the following:

Cysteine = cysteine hydrochloride — in final concentration of 1:500 for C203; 1:1,000 for pneumo.

P.A.S. = sulfanilamide — in final concentration of 1:10,000

S.P. = sulfapyridine — in final concentration of 1:10,000

‡ Jar = Brown anaerobic jar; seal = % in. vaseline; none = Wassermann tube, unsealed aerobic.



## SULFANILAMIDE AND ITS DERIVATIVES

prevailed in the medium. Later in the same year we (32) described the bacteriostatic effects of sulfanilamide in anaerobic cultures of *Cl. welchii*. This latter observation has been confirmed by Sadusk and Manahan (33).

Recently we have studied the bacteriostatic effects of sulfanilamide and sulfapyridine upon hemolytic streptococci and pneumococci under various conditions of anaerobiosis. These tests were conducted with rigid precautions to secure anaerobiosis and in each instance the reduction of methylene blue was used as an indication that anaerobiosis had been obtained. The results of these experiments are outlined in Table 2. The figures represent degrees of inhibition and not bacterial counts, hence the larger the figure the better the degree of bacteriostasis. It is evident from the data presented here that the bacteriostatic effects of sulfanilamide upon a strain of type I pneumococcus and upon the C203 strain of group A hemolytic streptococci were quite good under anaerobic conditions. When sulfapyridine was used as the bacteriostatic agent there was no difference between the effects noted under aerobic and anaerobic conditions in the pneumococcal cultures (in the majority of tests sulfapyridine sterilized the cultures irrespective of their oxygen content) and there was only a slight difference when the streptococcus was the test organism.

Regardless of the reason for the disagreement between our results and those of other workers who have studied the effect of anaerobiosis upon the bacteriostatic action of sulfanilamide, the fact that we obtained definite bacteriostasis with sulfanilamide and actual killing with sulfapyridine in the absence of air tends to discredit the concept that these drugs are activated by molecular oxygen. Obviously, if anaerobic conditions prevail, hydrogen peroxide cannot be formed from bacteria, as has been postulated. As a matter of record, our experience is in complete harmony with that of Shinn, Main, and Mellon (28), who found that at concentrations of less than 0.04 per cent of oxygen, sulfanilamide exerted a bacteriostatic effect in cultures of pneumococci.

## PERRIN H. LONG

Another observation which throws light upon the importance of peroxide in the mechanism of action of sulfanilamide was made by Fuller and Maxted (34), who noted that type III, group A hemolytic streptococci as a class fail to produce hydrogen peroxide. It just happens that the strains "Richards" and C203 are type III hemolytic streptococci and have been shown by Fuller and Maxted and ourselves not to produce detectable amounts of peroxide in cultures. Nevertheless both of these strains have been demonstrated to be very sensitive to the bacteriostatic activity of sulfanilamide. As Fuller and Maxted have put it, these observations constitute "an apparently insuperable objection" to the hypothesis of Locke and his associates.

The third theory dealing with the mode of action of sulfanilamide is based upon the idea that the drug acts on bacteria to prevent them from utilizing the substrate or upon the substrate to prevent it from being utilized by the bacteria. Levaditi (35), who was one of the first to champion this belief, considered that sulfanilamide was changed to a sulfur-protein complex in the body cells. This compound he called the "active principle X." He thought that it blocked "the assimilative potential of the nutritive materials which the body places at the disposal of the metabolism of the (virulent) streptococci."

Some time later Lockwood (36) independently arrived at a somewhat similar concept of the mode of action of sulfanilamide. His observation, that the addition of small amounts of peptone to human serum cultures definitely decreased the bactericidal and, in some instances, the bacteriostatic effects of the drug upon virulent hemolytic streptococci, led him to suggest "that sulfanilamide prevents the specialized metabolic activity required of invasive organisms" and "that this effect may be achieved through prevention of the utilization of the protein substrate by the organisms." Recently Lockwood (37) has noted a similar protective effect on the part of peptone against the action of sulfanilamide and sulfapyridine upon the pneumococcus in human serum and upon *E. coli* in urine cultures.

Weld and Mitchell (38) have taken certain exceptions to

## SULFANILAMIDE AND ITS DERIVATIVES

Lockwood's interpretation of his experimental data. These observers ask, "Is it not possible that the explanation, at least in part, of the greater effect of sulfanilamide on organisms growing in peptone-free serum than on organisms growing in serum containing peptone, is merely that the organisms multiply more slowly in the former than in the latter medium and hence sulfanilamide has fewer organisms, in the medium in which peptone is excluded, on which to act and therefore has a correspondingly greater effect?" They then point out that in rabbit serum streptococci multiplied rapidly at 37.5° C. and that with a relatively large inoculum (3,000 organisms) a 100 mg. per cent concentration of sulfanilamide had only a slight bacteriostatic effect regardless of whether or not the serum cultures contained peptone. If, however, the inoculum was small (400 organisms per ml.) sulfanilamide was bactericidal for organisms growing in peptone-free serum and "merely bacteriostatic for those growing in the medium containing peptone." Their conclusion, drawn from these observations, is that the better the serum medium is for bacterial growth, the less will be the effect of sulfanilamide.

These observers also pointed out that White and Parker (39) had previously noted that sulfanilamide was a much more effective bacteriostatic and bactericidal agent against streptococci when the tests were carried out at 40° C. To their surprise Weld and Mitchell found that sulfanilamide was apparently without bacteriostatic effect in tests run at 20° C., and if their tests were performed at 39° C., sulfanilamide was definitely more effective as a bacteriostatic agent in the serum-peptone cultures than in serum alone. They were at a loss to explain this observation, especially since they had noted that streptococci multiplied more rapidly in the serum-peptone cultures.

The observation that sulfanilamide is so much more effective as a bacteriostatic agent at from 39° to 40° C. was shown not to be the result solely of a poorer growth response by the organisms at these temperature levels. The explanation of this increase of activity is not clear, although Marshall (40) has

## PERRIN H. LONG

recently suggested that the only process which has so high a temperature coefficient is one which involves the denaturation of protein.

We have repeated Lockwood's experiments regarding the effect of the addition of peptone to human-serum cultures of streptococci and find ourselves in complete agreement in so far as the experimental data are concerned. While we agree in general with the view of Weld and Mitchell, that sulfanilamide becomes less effective with an increase in the nutritive value of a medium for a given organism, we cannot wholly accept their interpretation. We (15) have noted, in the course of performing bacteriostatic tests with a strain of group G hemolytic streptococci, that while 10 mg. per cent of sulfanilamide was actively bacteriostatic against the organisms in Neopeptone, beef-infusion dextrose broth, 500 mg. per cent of the drug had scarcely any inhibitory effect upon the growth of this organism when the tests were made in peptone dextrose water, although the reproduction rate of the organisms was essentially the same in both media. This brings up the point which we have repeatedly stressed, that it is unwise to place too great a reliance upon data concerning the mode of action of sulfanilamide which have been gained from tests made with one medium. The behavior of the drug may vary markedly when different media are employed.

Another problem connected with the mode of action of sulfanilamide and its derivatives is that which concerns the effect of these compounds upon the soluble toxins and the toxic products of various bacteria. Bosse (41) was probably the first to report that the presence of Prontosil decreased the hemolysin production of streptococci. This was confirmed by Levaditi and Vaisman (42), who in addition concluded that Prontosil and Neoprontosil prevented the *in vitro* action of streptococcal leucocidin.

However, other observers (43-45) have failed to demonstrate a neutralizing effect by sulfanilamide or Neoprontosil upon the soluble "toxins" of the streptococcus *in vitro*. Thus, while this

## SULFANILAMIDE AND ITS DERIVATIVES

phase of the question is subject to dispute, there can be little doubt that the findings of King and his associates (46) — which demonstrated that in a 1:1,000 concentration, Neoprontosil, while not bacteriostatic, did cause a lowering of the hemolytic index of streptococci grown in rabbit plasma rabbit red cell clots—are of great interest, especially so since sulfanilamide did not produce this effect.

In 1937 Levaditi and Vaisman (47) reported that the endotoxins of the gonococcus, meningococcus, and *B. aertrycke* could be neutralized in vivo by sulfanilamide and certain of its derivatives. Staphylococcal toxin was, on the other hand, unaffected by these drugs. These observations have been confirmed and extended by Carpenter and his associates (48-52), who have shown that not only is gonococcal toxin neutralized both in vitro and in vivo by sulfanilamide, but also the toxins of *Staphylococcus aureus*, *Cl. welchii*, and *Cl. tetani*. They have also demonstrated that the administration of Neoprontosil by the *oral* route (but not when injected parenterally) protected the majority of mice against lethal doses of the toxins of the gonococcus, the hemolytic streptococcus, *Staphylococcus aureus*, *Cl. welchii*, and *Cl. botulinus*.

We personally have had no experience with this phase of the mechanism of action of sulfanilamide, but it seems to us, in view of the clear-cut results obtained by Levaditi and by Carpenter and his associates, that one must accept their evidence that sulfanilamide and its derivatives possess the power to neutralize certain bacterial toxins and endotoxins both in vitro and in vivo.

In discussing the host factor we must at once point out that there is no experimental evidence that sulfanilamide or its derivatives *stimulate* the activity of the reticulo-endothelial system. The activity of the phagocytes or antibodies is generally a secondary (although in some instances a very important) effect of sulfanilamide therapy.

Domagk (3) remarked upon the essentially normal appearance of the leucocytes in peritoneal exudates from mice which

## PERRIN H. LONG

had been infected with streptococci and treated with Prontosil. However, it was Levaditi (4) who first pointed out that the phagocytosis of streptococci might play an important part in the recovery of mice infected with these organisms and treated with Prontosil. This observation has subsequently been confirmed by almost all those who have studied the effects of sulfanilamide therapy upon the cellular response in animals infected with virulent group A hemolytic streptococci.

During the past three years we have been very much interested in trying to determine what happens to the infecting microorganisms in animals that have been treated with sulfanilamide and its derivatives. Early in our studies (6) we noted that if the streptococcal peritonitis, which can be produced by the intraperitoneal injection of 1,000 M.L.D. of strain C203, was permitted to develop to a stage in which from one to five coccal units could be seen in an oil-immersion field of the stained peritoneal exudate, and then treatment with sulfanilamide was started, two phenomena were observed. The first was that within two to four hours the multiplication of the cocci in the peritoneal exudates of the treated animals was definitely retarded; in other words, the drug's bacteriostatic effect became apparent. Secondly, with bacteriostasis, an increase in phagocytosis over that already existing was noted. If treatment was continued, the number of extracellular cocci steadily decreased as phagocytosis increased, until a point was reached when the exudate was free from visible streptococci. In contradistinction to the observation that, although phagocytosis might be quite brisk in the untreated mice, the ingested cocci multiplied within the cells and frequently destroyed them, was the finding that in the treated mice the streptococci did not multiply within the cells and were soon digested by the phagocytes.

The chain of events which we have just described is that noted in mice which received a relatively small inoculum of hemolytic streptococci and in which the infection was allowed to progress naturally for a period of six to eight hours before treatment was started. If, instead of this, the mice are treated

# SULFANILAMIDE AND ITS DERIVATIVES

on the day before and again one hour before being infected, and the infecting dose is large (about five million organisms), although a moderate degree of bacteriostasis and some phagocytosis may be noted in smears from the peritoneal exudate, the streptococci multiply rapidly and the mice succumb in from twelve to sixteen hours. An example of such an experiment is shown in Table 3.

TABLE 3.—FATE OF HEMOLYTIC STREPTOCOCCI (STRAIN C203) INJECTED INTO THE PERITONEAE OF MICE WHICH HAD BEEN PRETREATED AND THEN WERE TREATED WITH SULFAPYRIDINE

(Treatment: 20 mg. per os the day before and 1 hr. before infection and then 4 times a day for 3 days)

Mouse *	Inoculum M.L.D.	Time after Inoculation									
		5 min.		2½ hrs.		5 hrs.		7½ hrs.		12 hrs.	
		C.O.	%	C.O.	%	C.O.	%	C.O.	%	C.O.	%
		I.F.†	Phag.‡	I.F.	Phag.	I.F.	Phag.	I.F.	Phag.	I.F.	Phag.
C1	2.8 mil.	0.3	0	95	2	∞	1	Dead			
C2	2.8 mil.	0.4	0	Bloody tap		∞	1	Dead			
C3	2.8 mil.	0.9	0	27	3	400	0	...	4	∞	0
C4	2.8 mil.	1.0	0	62	8	∞	1	Dead			
T1	2.8 mil.	0.7	0	85	4	200	14	600	12	∞	30§
T2	2.8 mil.	5.5	0	118	3	300	1	600	16	1,000	28§
T3	2.8 mil.	5.7	0	117	1	500	18	∞	8	Dead	
T4	2.8 mil.	1.2	0	28§	0	500	12	700	16	∞	16§

\* C = controls; T = treated mice.

† C.O.I.F. = cocci per oil-immersion field.

‡ % Phag. = percentage of phagocytic cells showing ingested streptococci.

§ Dead in 18 hours.

However, if the streptococci have been grown for two or three generations in 20 mg. per cent sulfanilamide or sulfapyridine blood broth before being injected into the pretreated and subsequently treated mice, bacteriostasis and phagocytosis of the cocci begin almost at once, and within a few hours the peritoneal exudate is clear of streptococci. In untreated control animals the "pretreated" streptococci quickly regain their natural characteristics and the animals succumb within ten to twenty-four hours. This experiment (detailed in Table 4), when taken in conjunction with the one outlined in Table 3, shows that the

TABLE 4. — FATE OF HEMOLYTIC STREPTOCOCCI (STRAIN C203) GROWN IN BLOOD BROTH CONTAINING 20 MG. PER CENT SULFAPYRIDINE AND THEN INJECTED INTO THE PERITONEAE OF MICE WHICH HAD BEEN PRETREATED AND THEN WERE TREATED WITH SULFAPYRIDINE  
(Treatment: 20 mg. per os the day before and 1 hr. before infection and then 4 times a day for 3 days)

Mouse*	Inoculum MLD.	Time after Inoculation											
		5 min.		2 hrs.		5 hrs.		7½ hrs		12 hrs.		24 hrs.	
		C.O. I.F.†	% Phag‡	C.O. I.F.	% Phag.	C.O. I.F.	% Phag.	C.O. I.F.	% Phag.	C.O. I.F.	% Phag.	C.O. I.F.	% Phag.
C1	..... 5 mil.	4.9	0	0.2	0	0.6	5	6.3	0	500	8	Dead	
C2	..... 5 mil.	3.9	0	27	0	500	6	∞	78	Dead			
C3	..... 5 mil.	2.4	0	18	0	400	5	∞	76	Dead			
C4	..... 5 mil.	5.7	0	Bloody tap		Bloody tap		Bloody tap		∞	710	Dead	
T1	..... 5 mil.	3.5	0	0.8	4	0	8	0	2	0	0	0	0
T2	..... 5 mil.	4.5	0	2.2	6	0	4	0	0	0.9	0	0.5	0
T3	..... 5 mil.	6.1	0	2.6	16	8.6	6	0	8	0	1	0	0
T4	..... 5 mil.	4.1	0	0.4	7	0	3	0	2	0	1	0	0

\* C = controls; T = treated mice.

† C.O.I.F. = cocci per oil-immersion field.

‡ % Phag. = percentage of phagocytic cells showing ingested streptococci.



## SULFANILAMIDE AND ITS DERIVATIVES

prior injection of sulfanilamide or sulfapyridine into mice does not activate the drug, and that pretreatment of the streptococci with the drug makes them immediately susceptible to the effects of the drug when they are injected into mice which have been treated.

If the mouse is deprived of its polymorphonuclear leucocytes, as can be done by the administration of benzene, adequate therapy with sulfanilamide is unavailing; the streptococci multiply slowly and the mouse dies in from three to four days. This indicates that bacteriostasis alone is insufficient and that the presence of the leucocytes is necessary if the infection is to be brought under control.

It might be asked: Do not other host factors, such as the development of antibodies, come into play in the recovery of these mice? As far as anyone has been able to show, they do not. Mice which have recovered from a hemolytic streptococcal infection as a result of sulfanilamide therapy are as susceptible to reinfection with the homologous organism as are normal mice. Hence it seems that in mice the bacteriostatic activity of the drug, coupled with the resulting phagocytosis, is the essential factor in recovery from group A hemolytic streptococcal infection. In human beings this factor is certainly of great importance, but inasmuch as studies of the immune responses of patients ill with hemolytic streptococcal infections and treated with sulfanilamide have not been reported, we cannot say with certainty that it represents the whole host factor.

A second type of host response is noted in mice infected with virulent pneumococci. We (53) have previously shown that both sulfanilamide and 4:4' diamino diphenyl sulfone exerted a bacteriostatic effect upon the growth of pneumococci in the peritoneal cavities of mice, and that despite this bacteriostasis little phagocytosis was noted. The pneumococci multiplied at a definitely slower rate in the treated animals than in the controls, but all of the mice eventually died. It was also shown that the lack of phagocytosis was not the result of damage to the phagocytic cells, for when conditions were made favorable by injecting

TABLE 5.—FATE OF TYPE I PNEUMOCOCCI (STRAIN SVI) FOLLOWING THEIR INJECTION INTO THE PERITONEAL OF MICE WHICH HAD BEEN PRETREATED AND THEN WERE TREATED WITH SULFAPYRIDINE

(Treatment: 20 mg. per os the day before and 1 hr. before infection and then 4 times a day for 3 days)

Mouse*	Inoculum M.L.D.	Time after Inoculation																	
		5 min.		2 hrs.		4 hrs.		6 hrs.		8 hrs.		11 hrs.		24 hrs.		30 hrs.		48 hrs.	
		C.O.	%	C.O.	%	C.O.	%	C.O.	%	C.O.	%	C.O.	%	C.O.	%	C.O.	%	C.O.	%
		I.F.† Phag.‡ I.F. Phag. I.F. Phag. I.F. Phag. I.F. Phag. I.F. Phag. I.F. Phag. I.F. Phag. I.F. Phag. I.F. Phag.																	
C1	.....	5.2 mil.	0.8	0	18	0	107	0	300	0	500	?	∞	0	Dead				
C2	.....	5.2 mil.	2	0	22	0	125	0	300	0	500	?	∞	?	Dead				
C3	.....	5.2 mil.	1.2	0	22	0	123	0	250	0	300	0	500	?	Dead				
C4	.....	5.2 mil.	1.4	0	16	0	114	0	300	0	?	?	∞	?	Dead				
T1	.....	5.2 mil.	0.8	0	17	0	94	0	65		Bloody tap		Bloody tap		Bloody tap	3	0	0.8	0
T2	.....	5.2 mil.	0.9	0	20	0	92	0	120	0	100	0	30	2	1	2	0	0	0
T3	.....	5.2 mil.	0.9	0	12	0	118	0	65	0	30	0	19	1	10	1	3	1	0
T4	.....	5.2 mil.	0.3	0	16	0	114	0	96	0	76	0	45	?	32	2	8	?	0

\* C = controls; T = treated mice.

† C.O.I.F. = cocci per oil-immersion field.

‡ % Phag. = percentage of phagocytic cells showing ingested pneumococci.

## SULFANILAMIDE AND ITS DERIVATIVES

type-specific antipneumococcal serum, a wave of phagocytosis was noted.

Buttle (54) had found that mice which survived after being infected with fairly large numbers of pneumococci and treated with the benzylidene derivative of diamino diphenyl sulfone were immune to subsequent infection with the homologous type of pneumococcus. Whitby (55) confirmed this observation (as we have also) when sulfapyridine was used in the treatment of experimental pneumococcal infections in mice.

It therefore seemed from these observations that the mechanism of recovery, through the agency of drug therapy, differed in experimental pneumococcal and streptococcal infections in mice.

Table 5 gives the results of experiments designed to throw light upon this question. As will be noted, if mice are pretreated with sulfapyridine and then infected with a highly virulent (but sulfapyridine-susceptible) strain of type I pneumococci, a definite multiplication of the cocci takes place for the first four to six hours. Then (provided the treatments are kept up) a change takes place and, with but a minor degree of phagocytosis, the number of cocci rapidly decreases until practically none can be found in the peritoneal exudate at the end of forty-eight hours. What happens to the pneumococci during this period of decrease is unknown. They may be killed by the drug, but for this assumption we have no evidence. All that we can say with certainty is that they are not engulfed by the phagocytes. We do know, however, that if treatment is discontinued too soon, although the peritoneal exudate is free of cocci, the infection will recur promptly and the mice will die. Hence it is necessary to continue treatment for about five days if permanent survivals are desired.

If, as with the streptococci, the pneumococci are grown for one or two generations in blood broth containing 20 mg. per cent of sulfapyridine and are then injected into mice which have been pretreated and are treated with sulfapyridine, as is shown in Table 6, bacteriostasis begins almost immediately and the cocci slowly disappear from the exudate. Here again the

TABLE 6.—FATE OF TYPE I PNEUMOCOCCI (STRAIN SVI) GROWN IN BLOOD BROTH CONTAINING 10 MG. PER CENT SULFAPYRIDINE AND THEN INJECTED INTO THE PERITONEAE OF MICE WHICH HAD BEEN PRETREATED AND THEN WERE TREATED WITH SULFAPYRIDINE  
(Treatment: 20 mg. per os the day before and 1 hr. before infection and then 4 times a day for 3 days)

Mouse*	Inoculum MLD.	Time after Inoculation											
		5 min.		2 hrs.		4 hrs.		6 hrs.		10 hrs.		24 hrs.	
		C.O.	%	C.O.	%	C.O.	%	C.O.	%	C.O.	%	C.O.	%
		I.F.† Phag.‡		I.F. Phag.		I.F. Phag.		I.F. Phag.		I.F. Phag.		I.F. Phag.	
C1	..... 3.1 ml.	4.4	0	12	0	18	0	16	0	17	0	∞	0
C2	..... 3.1 ml.	4	0	9	0	Bloody tap		84	0	200	0	Dead	
C3	..... 3.1 ml.	6.4	0	12	0	12	0	100	0	500	0	Dead	
C4	..... 3.1 ml.	7	0	12	0	16	0	22	0	26	0	∞	0
T1	..... 3.1 ml.	6.2	0	3.6	0	3	0	4	0	4	0	3	0
T2	..... 3.1 ml.	5	0	6	0	5.6	0	4	0	4	0	1.6	0
T3	..... 3.1 ml.	5	0	5.6	0	3.4	0	4	0	4.6	0	2	1
T4	..... 3.1 ml.	4	0	4.4	0	4.4	0	5	?	3.2	0	1.8	0
												2	0
												0	0.8

\* C = controls; T = treated mice.

† C.O.I.F. = cocci per oil-immersion field.

‡ % Phag. = percentage of phagocytic cells showing ingested pneumococci.

## SULFANILAMIDE AND ITS DERIVATIVES

phagocytes do not appear to play a prominent part in the removal of the pneumococci. Also, as in the previous experiment, if treatment is stopped when the exudate is just clear of pneumococci, a recurrence of the infection will almost invariably follow.

It seems quite clear in experimental pneumococcal infections that while the drug is holding the infecting organisms in check, they must be acting antigenically to bring about the type-specific immunization of the animal. Hence, if treatment is continued until the animal is well immunized, recovery from the infection is permanent. That this is probably true in human beings also is evident from the observations of Wood and Long (56), who noted in patients ill with pneumococcal lobar pneumonia and treated with sulfapyridine that although the therapy might bring about a rapid, apparent recovery, recurrences of the infection were to be expected if sulfapyridine was discontinued before type-specific antipneumococcal antibody appeared in their sera.

We may therefore say that the host response to chemotherapy in pneumococcal infections differs in both the experimental animal and the human being from that observed in hemolytic streptococcal infections.

A third type of host response has been observed in mice infected with *Cl. welchii* and treated with sulfanilamide. A typical protocol is shown in Table 7. It is to be noted in these experiments that without pretreatment of the organisms, but with the mice treated thirty minutes before being infected, the bacteriostatic effects of the drug are immediately evident. There is no lag period in the action of the drug in this type of infection. It is also of interest to observe that in the beginning phagocytosis is equal in both the control mice and the treated mice, but that as time goes on less phagocytosis is noted in the treated mice—obviously because there are fewer bacilli to be engulfed, owing to the inhibitory action of the drug upon the reproduction of the microorganisms. This, therefore, represents a host response in which no bar to phagocytosis exists in either the

TABLE 7.—EFFECT OF SULFANILAMIDE THERAPY ON THE PERITONEAL EXUDATE IN  
EXPERIMENTAL CLOSTRIDIUM WELCHII PERITONITIS IN MICE

Mouse	Inocu- lum	Total Therapy	Time after Infection															
			15 min.	2¼ hrs.	4¼ hrs.	6¼ hrs.	8¼ hrs.	24 hrs.										
			C.I.O.	C.I.O.	%	Cl.	C.I.O.	%	Cl.	C.I.O.	%	Cl.	C.I.O.	%	Cl.	C.I.O.	%	Cl.
			I.F.*	I.F. Phag.†	100	Phags.‡	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100	I.F. Phag. 100
1.....	0.5 cc.	28 mg.	21	6	60	269	0.5	36	92	1	17	44	1	30	69	0	22	Dead
2.....	0.5 cc.	28 mg.	29	13	84	498	1	39	160	4	30	140	1	27	84	0	14	Dead
3.....	0.5 cc.	28 mg.	12	13	98	502	15	48	204	2	15	89	1	28	102	0	17	Dead
4.....	0.5 cc.	Control	15	55	96	598	104	96	712		Dead							Dead
5.....	0.5 cc.	Control	18	96	98	898	109	100	1190		Dead							Dead
6.....	0.5 cc.	Control	31	360	98	822		Dead										

\* C.I.O.I.F. = number of free clostridia per oil-immersion field.

† % Phag. = percentage of phagocytes containing clostridia.

‡ Cl. Phag. 100 Phags. = number of clostridia ingested by 100 phagocytes.

## SULFANILAMIDE AND ITS DERIVATIVES

treated or control animals, and which clearly demonstrates the bacteriostatic activities of the drug *in vivo*.

It seems from the experiments just outlined that the nature of the host response is of importance if a clear picture of the mechanism of action of sulfanilamide and its derivatives in the control of infections is to be obtained. Thus far we have been able to demonstrate two main types of host response. In the first, phagocytosis seems to play an important role in finally ridding the animal of the infectious agent, while specific antibody production either does not occur or is of minor importance. In the second type of host response, primary phagocytosis is slight and the drug exercises merely an inhibitory effect upon the infectious agent until the naturally developing specific immune bodies are able to cope with the infection. A third type of host response may exist, in which, while phagocytosis is very important, the primary factor in recovery seems to be the immediate bacteriostatic effects of the drug upon the invading microorganisms.

*Summary and Conclusions.*—It seems quite probable from the evidence now available that the chemotherapeutic activity of Prontosil and Neoprontosil in streptococcal infections is the result of their breakdown to sulfanilamide in the tissues of the infected host.

Sulfanilamide and sulfapyridine act as bacteriostatic and, under certain conditions, as bactericidal agents against susceptible bacteria. They also seem to have the power of inactivating certain bacterial toxins.

The mechanism by which these drugs produce bacteriostasis and inactivation of toxins remains unknown despite several attempts to explain it.

The importance of the response of the host to an infection which is treated by chemotherapy should be recognized if a complete picture of the mode of action of these drugs is to be acquired.

## PERRIN H. LONG

### *References*

1. Shaffer, P. A. 1939. *Science* 89:547.
2. Foerster. 1938. *Zentralbl. f. Haut- u. Geschlechtskr.* 45:459.
3. Domagk, G. 1935. *Deutsche med. Wochenschr.* 61:250.
4. Levaditi, C., and Vaisman, A. 1935. *Compt. rend. de acad. de sci.* 200:1694.
5. Colebrook, L., and Kenny, M. 1936. *Lancet* 1:1279.
6. Long, P. H., and Bliss, E. A. 1937. *Jour. Am. Med. Assn.* 108:32.
7. Fuller, A. T. 1937. *Lancet* 1:194.
8. Levaditi, C., and Vaisman, A. 1935. *Compt. rend. de soc. de biol.* 119:946.
9. ——— 1935. *Presse méd.* 43:2097.
10. Domagk, G. 1936. *Klin. Wochenschr.* 15:1585.
11. ——— 1938. *Dermat. Wochenschr.* 107:797.
12. ——— 1937. *Zeitschr. f. klin. Med.* 132:775.
13. ——— 1939. *L'Europe méd.* Vol. 4, Nos. 3 and 4.
14. Tréfouël, Mme. J., Nitti, F., and Bovet, D. 1935. *Compt. rend. de soc. de biol.* 120:756.
15. Long, P. H., and Bliss, E. A. 1939. *The Clinical and Experimental Use of Sulfanilamide, Sulfapyridine and Allied Compounds.* The Macmillan Company, New York.
16. Fourneau, E., Tréfouël, J., Tréfouël, Mme. J., Nitti, F., and Bovet, D. 1936. *Compt. rend. de soc. de biol.* 122:652.
17. Colebrook, L., Buttle, G. A. H., and O'Meara, R. A. Q. 1936. *Lancet* 2:1323.
18. Bliss, E. A., and Long, P. H. 1937. *Bull. Johns Hopkins Hosp.* 60:149.
19. Mayer, R. L., and Oechslein, C. 1939. *Compt. rend. de soc. de biol.* 130:211.
20. Mayer, R. L. 1937. *Biol. méd. Suppl.* 27:45.
21. Ottenberg, R., and Fox, C. 1938. *Proc. Soc. Exper. Biol. Med.* 38:479.
22. Fox, C., Cline, J. E., and Ottenberg, R. 1939. *Jour. Pharmacol. Exper. Therap.* 66:99.
23. Fox, C., German, B., and Janeway, C. A. 1939. *Proc. Soc. Exper. Biol. Med.* 40:184.
24. Warren, J., Street, J. A., and Stokinger, H. E. 1939. *Proc. Soc. Exper. Biol. Med.* 40:208.
25. Locke, A., Main, E. R., and Mellon, R. R. 1938. *Jour. Immunol.* 36:183.
26. Main, E. R., Shinn, L. E., and Mellon, R. R. 1938. *Proc. Soc. Exper. Biol. Med.* 39:272.
27. Shinn, L. E., Main, E. R., and Mellon, R. R. 1938. *Proc. Soc. Exper. Biol. Med.* 39:591.
28. ——— 1939. *Proc. Soc. Exper. Biol. Med.* 40:640.
29. Locke, A., Main, E. R., and Mellon, R. R. 1938. *Science* 88:620.
30. Locke, A., and Mellon, R. R. 1939. *Science* 90:231.
31. Mellon, R. R., Locke, A., and Shinn, L. E. 1939. *Abstr. Third Internat. Congress Microbiol.* p. 255.
32. Long, P. H., and Bliss, E. A. 1937. *Can. Med. Assn. Jour.* 37:457.



## SULFANILAMIDE AND ITS DERIVATIVES

33. Sadusk, J. F., and Manahan, C. P. 1939. Jour. Am. Med. Assn. 113:14.
34. Fuller, A. T., and Maxted, W. R. 1939. Brit. Jour. Exper. Path. 20:177.
35. Levaditi, C. 1937. Monogr. de l'inst. A. Fournier. No. 5.
36. Lockwood, J. S. 1938. Jour. Immunol. 35:155.
37. ——— 1939. Abstr. Third Internat. Congress Microbiol. p. 253.
38. Weld, J. T., and Mitchell, L. C. 1939. Jour. Bact. 38:335.
39. White, H. J., and Parker, J. M. 1938. Jour. Bact. 36:481.
40. Marshall, E. K., Jr. Personal communication.
41. Bosse, O. A. 1935. Inaug. Dissert. Dok. Med. Marburg.
42. Levaditi, C., and Vaisman, A. 1935. Compt. rend. de soc. de biol. 120: 1077.
43. Garrod, L. P. 1938. Lancet 1:1125, 1178.
44. Kemp, H. A. 1938. Texas State Jour. Med. 34:208.
45. Madison, R. R., and Snow, J. E. 1937. Proc. Soc. Exper. Biol. Med. 36:592.
46. King, J. T., Henschel, A. F., and Green, B. S. 1938. Proc. Soc. Exper. Biol. Med. 38:810.
47. Levaditi, C., and Vaisman, A. 1937. Compt. rend. de acad. de sci. 205: 1108.
48. Carpenter, C. M., Hawley, P. L., and Barbour, G. M. 1938. Science 88:530.
49. Carpenter, C. M., Barbour, G. M., and Hawley, P. L. 1938. Jour. Bact. 36:280.
50. ——— 1939. Jour. Pediat. 14:116.
51. Carpenter, C. M., and Barbour, G. M. 1939. Proc. Soc. Exper. Biol. Med. 41:354.
52. Carpenter, C. M. 1939. Abstr. Third Internat. Congress Microbiol. p. 256.
53. Long, P. H., Bliss, E. A., and Feinstone, W. H. 1939. Jour. Am. Med. Assn. 112:115.
54. Buttle, G. A. H. 1937. Proc. Roy. Soc. Med. 31:154.
55. Whitby, L. E. H. 1938. Lancet 2:1095.
56. Wood, W. B., Jr., and Long, P. H. 1939. Ann. Int. Med. 13:612.

# CHEMISTRY IN URINARY ANTISEPSIS

BY

Henry F. Helmholz

*Professor of Pediatrics, The Mayo Foundation,  
University of Minnesota*

THE history of urinary antiseptics is of interest. It must be remembered that bacteriology was an unknown science before 1880, and it was not until much later that it was extensively used in determining the actual causative bacteria in infections of the urinary passages.

It is understandable that therapeutic efforts should first be made to change the chemical reaction of the urine from alkaline to acid, or vice versa, depending on the condition of the urine at the time infection was discovered. For instance, Bokai in 1878 (1) recommended the alkalization of the urine with sodium bicarbonate in infections with an acid urine, and in cases of ammoniacal cystitis the acidification of the urine with benzoic, salicylic, or carbolic acid. As long as no limits were known for the alkalinity or acidity of the urine and no studies had been carried out to determine the bactericidal range of the alkalinity or the acidity, very little definite information was available regarding the control of clinical studies. Only much later did numerous studies show that with the alkalinity possible to obtain in human urine, pH 8, it is impossible to kill off any of the bacteria usually found in urinary infections. It was found that of the various types of organisms isolated from urine, none was killed in less than twenty-four hours, at a pH of 9, and it was further found that *Staphylococcus* and the *Proteus* organisms grew well even at pH 9.5. It is evident therefore that whatever the beneficial effect of alkalization of the urine, it is not a bactericidal action on the infecting organism, nor is there any clinical evidence that alkalization acts beneficially on the mucous membrane of the urinary passages.

## CHEMISTRY IN URINARY ANTISEPSIS

By means of acidification of the urine very definite clinical results are possible. A  $pH$  of 4.8 can be quite readily attained by the administration of acids or acid salts. At this  $pH$  a considerable number of organisms will show a tendency to be inhibited in their growth; and some, such as *Proteus ammoniae*, will be killed off completely. It is very questionable, however, whether acid therapy before the introduction of ammonium and calcium salts ever reached bactericidal ranges. If the bactericidal range of 5–4.8 can be reached, the addition of one of the urinary antiseptics, even in low concentration, gives an almost immediate effect, so that it seems futile to rely on acidity alone.

Methenamine was first prepared by Butlerow in 1860. Twenty-four years later the first great step forward in urinary antiseptics was made by Nikolaier (2), when he showed that the splitting of methenamine in the urine had an antiseptic effect. Methenamine is an inert substance which can be administered by mouth or intravenously in high concentrations without any ill effects. As long as it remains in an alkaline medium it remains unchanged. When the medium becomes acid, the drug is split into ammonia and formaldehyde, the latter producing bactericidal action. The drug is excreted to a large extent by the kidney, so that when it passes from the alkaline plasma to an acid urine it immediately begins to break down. As Shohl and Deming have shown (3), the rapidity of the development of bactericidal power in the urine depends on the  $pH$  of the urine, on the concentration of the methenamine, and on the temperature. Formaldehyde is bactericidal in low concentrations, but it is also very irritating to the mucous membranes of the urinary tract.

In experiments which I carried out in 1931 (4) it was possible to show that the bactericidal power of urine containing methenamine varied with the concentration of the methenamine, the  $pH$  of the urine, and the length of time in which the action took place. I showed that urine with a  $pH$  of 6.0, to which 0.5 per cent of methenamine was added, usually showed marked increase in the number of bacteria at the end of twenty-four

## HENRY F. HELMHOLZ

hours; and that urine with the same concentration of methenamine and a  $pH$  of 5.5 was usually sterile after six hours, and at a  $pH$  of 5.0 after four hours. Urine having a concentration of 0.25 per cent of methenamine and a  $pH$  of 5.5 was sterile after twenty-four hours. When the concentration was only 0.05 per cent there was luxuriant growth.

It was shown that the chemical reaction necessary to bring about a bactericidal urine in four hours required a concentration of 0.5 per cent of methenamine and a  $pH$  of less than 5.5. By giving ammonium chloride by mouth the  $pH$  of the urine can be reduced to less than 5.5; by increasing doses of methenamine the concentration of the drug can be raised to 0.5 per cent or greater; by allowing the drug to remain in the urinary passages for four hours a successful therapeutic result is probable. The irritating effect of the formaldehyde on an acutely inflamed bladder often interferes with successful administration of methenamine. In successful methenamine therapy the bactericidal effect of the formaldehyde makes itself manifest before the irritation of the bladder necessitates discontinuance of medication; and we ought never to despair of successful treatment until, with the urine at a  $pH$  less than 5.5, the drug has been increased to a point where either the urine has been sterilized or irritability of the bladder precludes further increase of the drug. In patients with normal urinary passages treatment was successful in many cases, but failure was almost invariable in cases with obstruction in the urinary tract.

Phenyl salicylate was used in cases where methenamine was not successful because of irritation of the bladder or other causes. This drug was recommended in infections of the urinary tract because it was split into salicylic acid and phenol in the intestinal canal and supposedly excreted as such in the urine. We have administered as much as 5 grams of the drug a day to children without obtaining a bactericidal urine, although it turned dark rapidly after standing. It seems likely, therefore, that in the much smaller doses in which the drug is usually given, phenyl salicylate is of little or no value.

## CHEMISTRY IN URINARY ANTISEPSIS

The acriflavine group has been fairly extensively used in urinary infections. Its use was particularly recommended by Davis in 1924 (5). Experimentally, in rabbits, he was able to show a strongly bactericidal urine after its administration by mouth. His clinical results did not bear out this lead, however. In the acute group, where spontaneous cure is always difficult to rule out, thirteen out of eighteen cases were cured, and in the chronic only four out of twenty-seven. In our experience the drug in a dosage which repeatedly produced diarrhea and vomiting failed to sterilize the urine even temporarily.

Pyridium is a dye having remarkable antiseptic powers when dissolved in water, which it almost completely lost when a small amount of urine was added (6). Hexylresorcinol was introduced by Leonard as a urinary antiseptic with great possibilities. Used in alkaline solution it promised success in experiments in vitro. How extensively it is being used today is difficult to say, but judging from the literature it has been entirely replaced by other urinary antiseptics. I (7) used hexylresorcinol in the treatment of a series of urinary infections with little or no success; many of the infections that could not be cleared up with hexylresorcinol yielded promptly to treatment with methenamine.

As a result of much clinical experience it is the opinion of most observers that of the urinary antiseptics used before the ketogenic diet was introduced methenamine was the most useful drug. When properly controlled its use is followed by the clearing up of many urinary infections.

In all the previous work the chemist had produced the substance and the clinician tried out its therapeutic properties. With the use of the ketogenic diet this process was reversed; the clinicians introduced it and the chemist explained its action. A chance observation on the keeping qualities of the urine from a patient who was on the ketogenic diet for the treatment of epilepsy opened up an entirely new field in the treatment of urinary infection. In vitro experiments showed that the urine of a patient on the ketogenic diet was bactericidal for practi-

## HENRY F. HELMHOLZ

cally all bacteria found in infections of the urinary tract when the  $pH$  was less than 5.5 (8). It was thus shown that the body by its own metabolic processes could produce a strongly bactericidal urine. Clinical applications of these studies by Clark (9) in adults and by Helmholtz (8) in children showed the ketogenic diet to be superior to any mode of treatment that had been available up to that time, and even at present it is probably the one least likely to do damage to the urinary tract.

It remained for Fuller (10) to show that beta-oxybutyric acid was the ketone body responsible for the bactericidal action. Later Osterberg and Helmholtz (11) showed that bactericidal power of the urine depended on the concentration of beta-oxybutyric acid and the  $pH$  of the urine. The lower the concentration of the acid the lower the  $pH$  had to be, and the higher the concentration of acid the higher the  $pH$  of the urine could be in order to have a bactericidal urine. Again it was found that with the usual concentration of beta-oxybutyric acid attained by the ketogenic diet, a  $pH$  of 5.5 was necessary in order to obtain a bactericidal urine. In children it was possible to clear up severe infection of the kidney in cases where this accompanied hydronephrosis. As shown by Rosenheim (12), the relationship of  $pH$  and concentration of the organic acid to bactericidal function is probably the amount of free undissociated acid available at the particular concentration and  $pH$ . The higher the  $pH$  the greater the concentration of the acid necessary to make available the same amount of free undissociated acid. At a  $pH$  of 5.2, 16 per cent of the beta-oxybutyric acid is in the free undissociated form; at  $pH$  5.5, 8.7 per cent; and at  $pH$  5.7 only 5.7 per cent. The in vitro studies in ketonurine showed that with a concentration of more than 0.5 per cent and a  $pH$  less than 5.5 bactericidal urine could be produced. We have produced ketosis with concentrations of beta-oxybutyric acid as high as 2.5 per cent which should yield a bactericidal urine at a  $pH$  of 6.0.

Unfortunately the diet necessary to produce a high concentration of beta-oxybutyric acid is rather difficult to prepare and

## CHEMISTRY IN URINARY ANTISEPSIS

disagreeable to take because of its high fat content, and many patients, especially those in the acute stages of the disease, are unable to take it. Vomiting was a frequent symptom in adults even when only a part of the maintenance diet was given.

The principle was established that an organic acid could be excreted in sufficient concentration by the body to be bactericidal for practically all organisms found in the urinary infections at a *pH* that could easily be reached in human urine. The administration of beta-oxybutyric acid as such by mouth was very disagreeable and in the absence of ketosis the drug was destroyed in the intermediary metabolism.

It was a great advance, therefore, when Rosenheim (13) suggested the use of mandelic acid in place of the beta-oxybutyric acid. Its calcium salt is quite palatable and only rarely produces untoward gastric symptoms. The calcium, as well as the ammonium salt, produces the necessary acidity in the urine, and by adjusting the dose a proper concentration of the free acid can be obtained in the urine. The conditions necessary for its action are apparently exactly the same as those for beta-oxybutyric acid. The lower the *pH* the lower the concentration of the acid necessary to produce a bactericidal urine. Rosenheim has shown that the concentrations of the drug and the corresponding *pH* which are necessary for bactericidal action, as determined by Osterberg and Helmholtz (14), correspond exactly to their level of free undissociated acid. Thus only 0.25 per cent of the acid is necessary for bactericidal action at *pH* 5.0, while 2.4 per cent acid is necessary at *pH* 6, 7.8 per cent at *pH* 6.5, and 24 per cent at *pH* 7.0 in order to obtain the required amount of free undissociated acid. Intense mandelic acid therapy occasionally produces hematuria, but no other untoward symptoms.

It will be noted that methenamine, the ketogenic diet, and mandelic acid all require a definite acidity for the development of their actions. The necessary level seems to be at or about *pH* 5.5. The attainment of this level is the stumbling block in the treatment of many cases of urinary infections. Two distinct groups are here to be recognized: the group in which the kidney,

## HENRY F. HELMHOLZ

because of reduced function, cannot excrete the drug in sufficiently high concentration or a urine of low  $pH$  or both; and the group in which the infecting organism produces an alkalinity of the urine which persists in spite of the administration of acid and acid-forming salts. This deficiency is of particular significance when we realize the frequency with which unilateral involvement occurs in kidneys damaged by obstruction or infection or both. Even when the combined bladder urine reaches the proper degree of acidity and concentration of the drug, the affected kidney may be excreting a urine of high  $pH$  and low concentration of the drug. An organic acid which in the same concentrations would give a bactericidal urine at a higher  $pH$  would be very desirable.

Very shortly after the introduction of mandelic acid as a urinary antiseptic, Unshelm (15) used prontosil in the treatment of urinary infections and reported that it had no effect *in vitro*. The effective portion of this drug was found to be sulfanilamide. In the urine it appears in the free and conjugated forms of which only the former is bactericidal. Sulfanilamide does not require special conditions in the urine; it acts in both alkaline and acid urine, and, in low concentrations, even better in alkaline than in acid urine; it is excreted by the damaged kidney in bactericidal concentration. Sulfanilamide thus fitted right into the weakest spot of our armamentarium in the treatment of infections of the urinary tract. In a series of cases where the kidney was damaged, as evidenced by a blood urea of more than 50 mg. per 100 cc. of whole blood, it was possible to sterilize the urine permanently in some and temporarily in others, a result which we have been unable to accomplish with any other urinary antiseptic (16). The action of sulfanilamide has been found to be bactericidal in concentrations easily obtained in the urine.

As Long and Bliss have pointed out (17), there is still some difference of opinion regarding the level at which bactericidal action occurs. In our work, even at levels as low as 25 mg. per 100 cc., we have observed bactericidal effect on some strains of the species commonly found in infections of the urinary tract.



## CHEMISTRY IN URINARY ANTISEPSIS

In concentrations of 80 to 100 mg. per 100 cc. we find that about 80 per cent of the various strains of the different organisms are killed off by sulfanilamide. There remain, however, a few strains of *Escherichia coli*, *Aerobacter aerogenes*, *Proteus ammoniae*, *Pseudomonas*, and *Staphylococcus* which do not yield to sulfanilamide even at a level of 200 mg. per 100 cc. We have found this to be true in repeated experiments and are unable to explain this very wide range of bactericidal action in different strains of the same species. Most striking of all is the complete failure of sulfanilamide, even in much higher concentrations, to affect the growth of any strain of *Streptococcus faecalis*.

A bactericidal action on such different organisms as *Streptococcus haemolyticus*, *Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas*, and *Proteus ammoniae*, combined with a lack of action on *Streptococcus faecalis*, presents a very interesting chemical problem. The great variation in the concentration of sulfanilamide necessary to kill off different strains of the same species of gram-negative bacilli, streptococci, and staphylococci prompted us to determine whether this fastness to sulfanilamide held also for other urinary antiseptics.

Dr. Sickler in a recent study (18) found that certain strains were killed readily at low concentration of sulfanilamide and that certain other strains were very resistant to much higher concentration. Using Sickler's resistant and a susceptible strain of *Staphylococcus aureus*, *Escherichia coli*, *Aerobacter aerogenes*, *Proteus ammoniae*, and *Pseudomonas aeruginosa*, we again tested them in the same way after an interval of four months and found that our results corresponded very closely to those of Sickler (see Table 1).

The different strains of microorganisms were then tested for their resistance to sulfapyridine, mandelic acid, and beta-oxybutyric acid. The concentrations of the drug and the pH of the urine were chosen so as to obtain the growth of some species of bacteria and the death of others. Sulfapyridine was used in a concentration of 150 mg. per 100 cc. of urine with a pH of

# HENRY F. HELMHOLZ

TABLE 1. — DIFFERENTIAL BACTERICIDAL EFFECT OF VARIOUS URINARY ANTISEPTICS

Microorganism	Sulfa- nilamide 80 mg. % pH 7.6	Sulfa- pyridine 150 mg. % pH 7.7	Mandelic Acid 0.2% pH 5.22	Beta- oxybutyric Acid 2.47% pH 5.98	Benzo- chrome 0.2 gm. 4 times a day
STRAINS RESISTANT TO SULFANILAMIDE *					
<i>Streptococcus faecalis</i> ....	0	0	+	0	+++
<i>Staphylococcus aureus</i> ....	0	0	+	+++	+++
<i>Escherichia coli</i> .....	0	0	+	+	0
<i>Aerobacter aerogenes</i> ....	0	0	0	++	0
<i>Proteus ammoniae</i> .....	0	+++	++	0	0
<i>Pseudomonas</i> .....	0	0	0	+++	0
STRAINS SUSCEPTIBLE TO SULFANILAMIDE					
<i>Staphylococcus aureus</i> ....	+++	+++	+	0	+++
<i>Escherichia coli</i> .....	+	+++	++	+	0
<i>Aerobacter aerogenes</i> ....	++	++	0	0	0
<i>Proteus ammoniae</i> .....	+	+++	+++	0	0
<i>Pseudomonas</i> .....	+	+	0	++	0

\* Microorganisms studied by Dr. Sickler.

7.7. The corresponding figures for mandelic acid were 200 mg. per 100 cc. with a pH of 5.22, and for beta-oxybutyric acid they were 2.47 per cent with a pH of 5.98. Experiments were also carried on with benzochrome, a substance now under investigation. So far its concentration in urine cannot be determined accurately. The results given in the table were obtained with the urine of a five-year-old child that had received 200 mg. four times a day.

A very interesting differential bactericidal picture developed from these studies (see Table 1). The strains resistant and the strains susceptible to sulfanilamide were also resistant and susceptible to sulfapyridine, with a single exception. The effects of the two drugs were thus very nearly the same on different strains of the same bacterium. As seen in Table 1, *Streptococcus faecalis* grew luxuriantly in both the sulfanilamide and the sulfapyridine solution, but it was placed in the chart to show its reaction to the other antiseptics.

Tested against mandelic acid and benzochrome the specific

## CHEMISTRY IN URINARY ANTISEPSIS

resistance of individual strains of the different bacteria was completely lost. Here resistance appeared to be specific for the species. Tested with beta-oxybutyric acid the strains resistant to sulfanilamide were more easily killed than those susceptible to sulfanilamide. The bactericidal effect of benzochrome on *Streptococcus faecalis* and *Staphylococcus aureus* was in striking contrast to its total absence of effect on the gram-negative bacilli. At higher concentration of the drugs all the organisms were affected.

The whole series of new sulfone compounds that are being developed will have to be tested for their toxicity to animals and man, and their bactericidal effect on the bacteria will have to be found in urinary infections in the hope that an even better urinary antiseptic may be discovered.

The function of urinary antiseptics is to render the urine bactericidal so that the infected lining of the urinary passage can give forth its bacterial content completely and will not become reinfected in the process. It will in this way gradually become sterile, a process which may take weeks or months. Properly administered, the urinary antiseptics at our disposal should cure most of the infections in patients with intact kidneys and unobstructed urinary tracts. The greatest problem today lies in the treatment of patients with damaged kidneys and those with an alkaline urine produced by bacteria-splitting urea.

### *References*

1. Bokai, J. 1878. In C. Gerhardt's *Handbuch der Kinderkrankheiten*. Vol. 4, pt. 3, pp. 529-31.
2. Nikolaier, A. 1895. *Deutsche med. Wochenschr.* 32:897.
3. Shohl, A. T., and Deming, C. L. 1920. Hexamethylenamin. Its quantitative factors in therapy. *Jour. Urol.* 4:419-37.
4. Helmholz, H. F. 1932. The effectiveness of methenamine as a urinary antiseptic at various hydrogen-ion concentrations. *Jour. Pediat.* 1:73-81.
5. Davis, E. D. 1924. Urinary antiseptics: Clinical results following the oral administration of acriflavine. *Jour. Urol.* 11:29-44.
6. Gillespie, J. B. 1933. Experiments in the antibactericidal properties of pyridium and serenium. *Am. Jour. Dis. Child.* 45:254-70.

## HENRY F. HELMHOLZ

7. Helmholz, H.F. 1926. The therapeutic value of hexylresorcinol in chronic pyelitis in childhood. *Am. Jour. Dis. Child.* 32:396-400.
8. ——— 1931. The ketogenic diet in the treatment of pyuria of children with anomalies of the urinary tract. *Proc. Staff Meetings of Mayo Clinic* 6:609-13.
9. Clark, A. L. 1931. *Escherichia coli* bacilluria under ketogenic treatment. *Proc. Staff Meetings of Mayo Clinic* 6:605-08.
10. Fuller, A. T. 1933. The ketogenic diet: The nature of the bactericidal agent. *Lancet* 1:855-56.
11. Helmholz, H. F., and Osterberg, A. E. 1936. The bactericidal effect of levorotatory and racemic beta-oxybutyric acid in urine. *Jour. Urol.* 35:86-92.
12. Rosenheim, M. L. Personal communication.
13. ——— 1935. Mandelic acid in the treatment of urinary infections. *Lancet* 1:1032-37.
14. Helmholz, H. F., and Osterberg, A. E. 1936. Rate of excretion and bactericidal power of mandelic acid in the urine. *Jour. Am. Med. Assn.* 107:1794-96.
15. Unshelm, E. 1936. Zur Behandlung der Kindlichen Pyurie. *Arch. f. Kinderh.* 109:65-84.
16. Helmholz, H. F. 1938. Urinary infections in infancy and childhood. *Jour. Am. Med. Assn.* 111:1719-22.
17. Long, P. H., and Bliss, E. A. 1939. *The Clinical and Experimental Use of Sulfanilamide and Sulfapyridine and Allied Compounds.* Macmillan, New York, p. 203.
18. Sickler, J. R. In press. A study of the bactericidal activity of sulfanilamide in urine at various levels of hydrogen ion concentration. *Jour. Urol.*

**Part IV. Some Approaches to the Nervous  
Control of the Organism**



# THE PHYSICOCHEMICAL APPROACH TO THE MECHANISMS OF CONVULSIVE REACTIVITY

BY

Irvine McQuarrie

*Professor of Pediatrics, University of Minnesota*

BECAUSE of the human element ever present in the problems of clinical medicine the investigator in this broad field feels literally compelled to attempt their solution by the most direct and practical means available. Once he has discovered a workable, albeit a superficial, understanding of the etiology or pathogenesis of a given disease or an effective method for its prevention or treatment, the nature of his responsibility immediately constrains him to redirect his efforts toward some entirely new or more urgent problem instead of centering them upon the fundamental aspects of that already in hand. As a result medical science, in spite of its phenomenal development, still remains amazingly incomplete as regards the basic mechanisms involved in even the most commonplace phenomena.

There is no better example of such deficiency than that in our basic knowledge of the mechanism of convulsive reactivity. While the primary causes of most of the common convulsive disorders are already known, the nature of the physicochemical changes occurring in the brain just prior to and during a convulsion are still obscure. In the case of genuine or so-called idiopathic epilepsy not even the primary cause is known, to say nothing of the convulsive mechanism. No organic basis for this bizarre disorder has been discovered in spite of decades of careful searching by competent pathologists. All efforts to find evidence of some predisposing disturbance of the metabolism in epilepsy, analagous to the hypocalcemia of tetania parathyreoprivia or the hypoglycemia of hyperinsulinism, have likewise failed. For these reasons it now appears certain that further

progress toward a solution of the enigma of epilepsy must depend upon a clearer elucidation of the convulsive mechanism per se. The present report deals with experimental investigations designed to attack the problem from this viewpoint. Data from studies on the convulsive mechanisms in epilepsy, hypoparathyroidism, and hyperinsulinism are briefly reviewed.

*Factors involved in the convulsive reactivity in epilepsy.*—The most significant advance thus far achieved toward an ultimate solution of the riddle involved in the chronic convulsive states was the recent discovery, by means of the electroencephalograph, that the spontaneously occurring grand mal seizures of epilepsy, as well as convulsions induced experimentally, result from, or are intimately associated with, an abnormal "storm" of electrical discharges in the cerebral cortex, giving rise to characteristic changes in the so-called brain wave records (1). What remains to be determined in the case of epilepsy is not only the nature of the physicochemical reactions immediately associated with the development of such abnormal electrical activity but also the character of the innate or constitutional defect in the brain cells themselves, which may be assumed to be responsible for the periodic, spontaneous breakdown in the physiological mechanism for controlling the normal rhythm of electrical potentials. Unfortunately the only course of action open to the investigator is an indirect one. That is, the nature of the cerebral deficiency must be inferred from the results of in vivo experiments designed to determine the specific conditions under which convulsive phenomena occur. Although many studies of this type have been made, only a few representative experiments can be referred to in the present communication.

A variety of heterogeneous factors are known to be conducive to the occurrence of convulsions in the epileptic subject. Alkalosis, undue cerebral irritation from any cause, "superhydration" under certain conditions to be considered later, and a number of other factors are listed as inciting or contributing causes. In seeking a common mechanism for explaining the



## THE MECHANISMS OF CONVULSIVE REACTIVITY

convulsive effect of all of these diverse factors, a number of workers have stressed the importance of cerebral anoxia. Others have offered purely circulatory or mechanical explanations. The anoxic theory was based in part upon the finding of Lennox and Cobb (2) that an increased tendency to petit mal seizures resulted from breathing an atmosphere with reduced oxygen content. The ease with which seizures could be induced by voluntary hyperventilation of the lungs was found by these authors to be increased greatly when the oxygen tension of the air was lowered. However, Lennox and Gibbs (3) and others have failed to find any evidence of spontaneously occurring anoxemia or of a disturbance in cerebral blood flow in epileptic subjects. On the basis of more recent studies, Gibbs, Lennox, and Gibbs (4) have concluded that a reduction in the carbon dioxide content of the plasma bears a more direct relationship to the occurrence of petit mal attacks than does lowered oxygen tension. From similar studies on patients suffering from grand mal, they have reported that the carbon dioxide content of the plasma increases just prior to the onset of seizures. Teglbjaerg (5) was unable to demonstrate any increase in convulsive tendency in epileptic patients as a result of lowered oxygen tension. Objections to the circulatory and mechanical theories have been discussed by Cobb (6).

On the basis of the present-day conception that various stimuli induce excitation by altering the concentrations of ions on cell membrane surfaces, a number of workers in this field (7, 8, 9, 10) have more or less independently advanced the hypothesis that generalized convulsions are dependent upon a profound disturbance in the surface functions of the cerebral cortical cells. Presumptive evidence of some such dysfunction is found in the observation that the various factors known to increase the convulsive tendency happen to be those usually regarded as effecting an increase in cell membrane permeability, at least under the conditions prevailing in simple biological experiments (see Table 1). More direct evidence in favor of this view has recently been furnished by the studies of Spiegel and

# IRVINE McQUARRIE

Spiegel-Adolf (11), in which the permeability of cerebral tissue for ions was determined indirectly on the intact brain (cat and man) by a special electrical conductivity method. Epileptogenous agents or procedures were found to increase brain cell membrane permeability, as estimated by this means, while anti-convulsants, such as sedatives, anesthetics, and hypnotics, were found to exert the opposite effect.

TABLE 1. — SUGGESTIVE PARALLELISM BETWEEN CONVULSIVE REACTIVITY AND PERMEABILITY OF BRAIN CELL MEMBRANES IN EPILEPSY, AS INFLUENCED BY VARIOUS PHYSIOLOGICAL FACTORS

Physiological Factor	Tendency to Convulsions	Membrane Permeability
Anesthesia	Decreased	Decreased
Sedation		
Acidosis		
Ketosis		
Dehydration		
Low $\frac{\text{lecithin}}{\text{cholesterol}}$ of serum	Increased	Increased
Excitation		
Brain trauma		
Alkalosis		
Superhydration		
Anoxemia		
Hypocalcemia		
High $\frac{\text{lecithin}}{\text{cholesterol}}$ of serum		

In the course of an experimental study on the relationship of water balance to the occurrence of convulsions in epilepsy, the author and Peeler (12) obtained results which were interpreted as indirect evidence of some inherent defect in the brain cell membranes of patients suffering from this disorder. The investigation referred to showed that typical convulsions could be induced practically at will in epileptic patients, but not in non-epileptic control subjects, by administering antidiuretic, post-pituitary extract (pitressin) at three-hour intervals for from twelve to thirty-six hours, during periods when the mineral intake was extremely low while the water intake was unrestricted. Under these special conditions the retention of a comparatively

## THE MECHANISMS OF CONVULSIVE REACTIVITY

large volume of water (equivalent to from 2 to 5 per cent of the original body weight), without a corresponding retention of electrolytes, results in a significant dilution of the extracellular body fluids. That this dilution is essential for the induction of the epileptic convulsion under these circumstances is demonstrated by the fact that seizures do not occur if dilution is prevented by administration of a sufficient amount of sodium chloride to make the fluid retained approximately isotonic (see

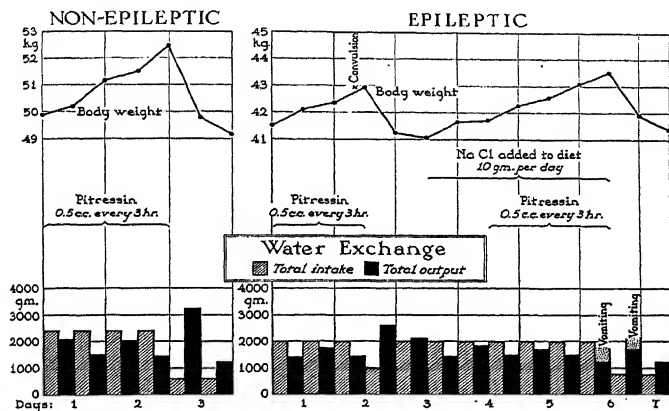


Figure 1. Characteristic convulsive response of patient with idiopathic epilepsy to sustained postpituitary antidiuresis with high water, low mineral intake. Also effect of preventing dilution of extracellular body fluids. (Drawn in the Medical Art Shop, University of Minnesota.)

Figure 1). As regards the convulsive mechanism, it may be assumed that dilution of the extracellular body fluids results in temporary passage of extra water into the cells of the brain in connection with the process of restoring osmotic equilibrium. That potassium may "leak" from the cells under such conditions was suggested by the results of another study (10), in which a strongly negative potassium balance developed during the period of sustained antidiuresis. This phenomenon was in evidence even before seizures began to occur. That other inor-

ganic ions ( $\text{Cl}^-$ ,  $\text{Na}^{++}$ ,  $\text{HPO}_4^-$ , etc.) also migrate across the cell membranes more readily than usual under such conditions is quite possible, although proof for such translocation of elements is entirely lacking at the present time. Confirmatory evidence regarding the effect of dilution of body fluids on the occurrence of convulsions in epilepsy was recently obtained by Ziskind and Somerfeld-Ziskind (13), who found that seizures could be induced practically at will in a majority of epileptic subjects merely by having them drink water at the rate of five to seven liters daily.

On the tentative assumption that the central physicochemical reactions associated with a convulsion, like its outward manifestations, may be essentially the same (no matter what the primary or inciting cause), investigations on the convulsive mechanism in hypoparathyroidism and in hyperinsulinism were carried out with the hope that they might yield information of value in the final solution of the more perplexing problem of epilepsy. Representative data from these more recent studies are presented below.

*Factors concerned in the convulsive mechanism in hypoparathyroidism.*—The lack of agreement, not infrequently observed, between the level of serum calcium and the occurrence of generalized convulsions in the "low-calcium" types of tetany indicates the need for further information regarding other factors involved in the convulsive mechanism in such disorders. That a profound disturbance in the metabolism of calcium and phosphorus is primarily responsible for the initial development of the neuromuscular hyperexcitability characterizing these disorders cannot be gainsaid, however. In these particular forms of tetany neither the milder symptoms nor the generalized convulsions occur without a marked decrease in the calcium of the blood plasma and other extracellular body fluids. Furthermore, all available evidence indicates the special significance of that portion of the calcium occurring in ionized form, for it alone appears to play a role in restraining neuromuscular hyperexcitability.

## THE MECHANISMS OF CONVULSIVE REACTIVITY

That decreased  $\text{Ca}^{++}$  concentration does not alone afford an adequate explanation for the various manifestations of active tetany, particularly the convulsive tendency, is apparent, however, from a number of considerations. For instance, it has repeatedly been observed that parathyroidectomized animals, showing signs of active tetany and convulsions during the early postoperative days, may later become entirely symptom-free by some unknown mechanism of adjustment (14, 15), not necessarily involving changes in the serum calcium, inorganic phosphorus, protein, or hydrogen-ion concentration. As abundantly illustrated in the present study, patients with idiopathic hypoparathyroidism not infrequently show periods of comparative freedom from convulsions without corresponding chemical changes in the blood. Acute infections tend to precipitate convulsive seizures in subjects with latent tetany without necessarily affecting the  $\text{Ca}^{++}$  concentration of the serum. While the signs of active tetany can be induced in subjects who are in a state of latent tetany by certain procedures known to decrease  $\text{Ca}^{++}$  concentration, such as alkalosis resulting from hyperventilation of the lungs or excessive administration of alkaline salts, attempts to demonstrate a significant swing in the acid-base equilibrium of the body fluids toward the alkaline side just prior to the onset of spontaneous convulsions have so far failed. Alkalosis per se is known to increase neuromuscular hyperexcitability apart from its effect on the  $\text{Ca}^{++}$  concentration. By means of direct determination of the  $\text{Ca}^{++}$  concentration (biological method) McLean, Barnes, and Hastings (16) demonstrated that this factor is only one of the variables concerned. They showed, too, by this direct method that other factors, such as the acid-base balance, known to influence the activity of tetany, do not act solely by influencing the ionization of calcium.

The foregoing observations suffice to emphasize the importance of obtaining more experimental data pertaining to other factors that contribute to the occurrence of convulsions in hypoparathyroidism. With this in mind experiments were carried out in collaboration with Dr. A. E. Hansen and Dr. M. R.

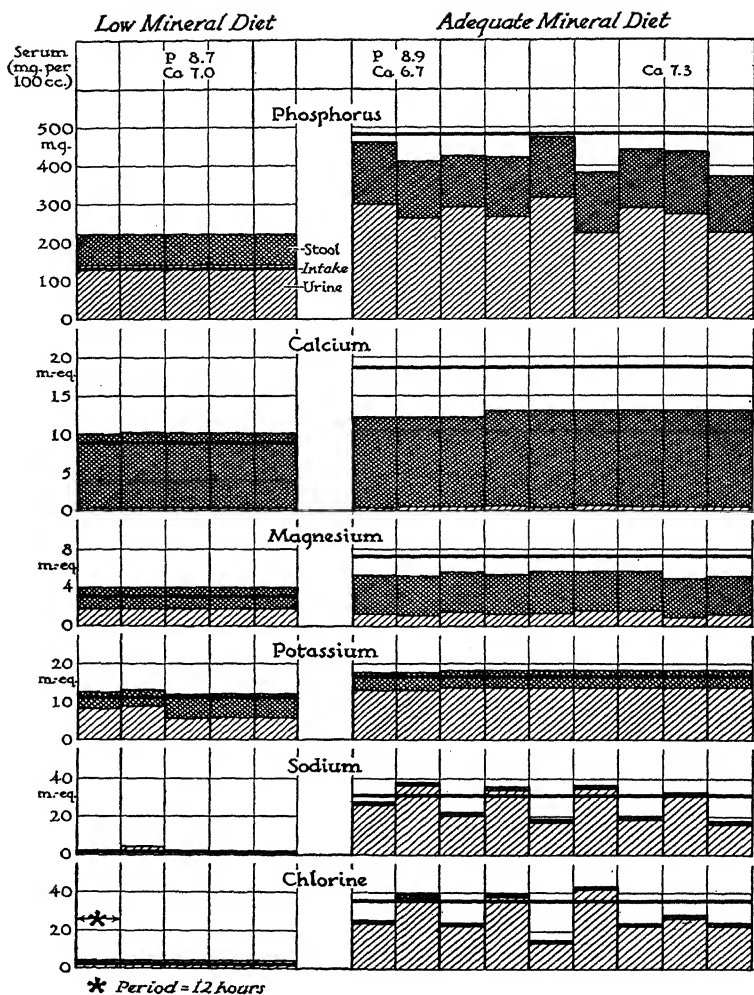


Figure 2. Hypoparathyroidism. Patient L. W., age 15 years. Representative data on mineral exchanges during periods of low and of high or adequate mineral intake. (Drawn in the Medical Art Shop, University of Minnesota.)

## THE MECHANISMS OF CONVULSIVE REACTIVITY

Ziegler in the case of a fifteen-year-old boy who had suffered for several years from idiopathic parathyreoprivic tetany in typical form. The general aim of the investigation was to ascertain the degree of correlation between changes in the serum electrolytes and in the water and mineral balances on the one hand and the occurrence of convulsions on the other. The effects of a variety of experimental procedures or conditions, such as changes in the mineral content of the diet, variation in the water intake, "forced" dilution of the extracellular body fluids (sustained, post-pituitary antidiuresis), voluntary hyperventilation of the lungs, administration of parathyroid gland extract, intravenous administration of typhoid vaccine (simulating acute bacterial infection) and artificial fever induced by diathermy, were determined. The first series of acute experiments were carried out after the subject had been maintained for a number of weeks on a diet characterized by its extremely low mineral content. A second series consisted of similar experiments repeated some weeks after he had been adjusted to a diet whose mineral content was comparatively high and considered to be optimal for normal growth. Samples of data from mineral balance studies representative of both periods are shown in Figure 2.

The results of the various experiments may be summarized as follows (see Figure 3): It was found that convulsions could be induced readily by voluntary hyperventilation of the lungs when the patient was on the low-mineral diet, but could not be induced at all, or only after considerably longer periods of hyperventilation, when he had subsisted for some weeks on the diet with a higher mineral content. Likewise, typical, generalized convulsions could be induced regularly during periods of low mineral intake by sustained post-pituitary (pitressin) antidiuresis, which caused retention of a comparatively large volume of water without a corresponding retention of electrolytes (Figure 3, a). During a subsequent low-mineral period it was found that the convulsive tendency produced by this latter procedure was greatly reduced when an amount of sodium chloride calculated to be sufficient for preventing dilution of the extra-

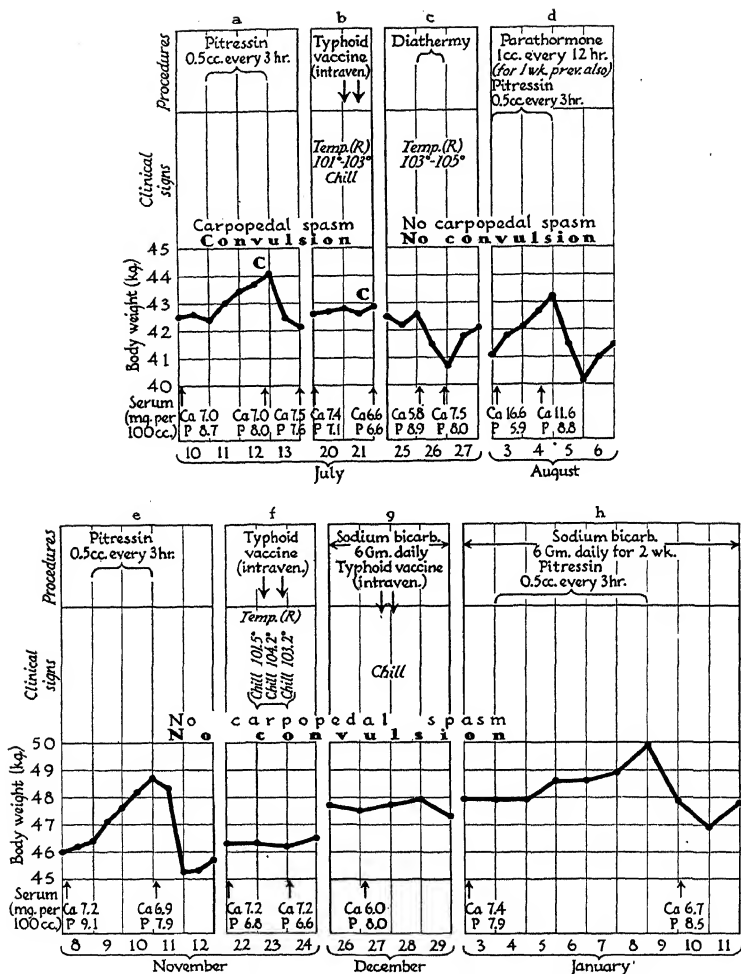


Figure 3. Hypoparathyroidism. Patient L. W., age 15 years. Effects of various experimental procedures during period of low mineral intake (a, b, c, d) and during period of high mineral intake (e, f, g, h). See text. (Drawn in the Medical Art Shop, University of Minnesota.)



## THE MECHANISMS OF CONVULSIVE REACTIVITY

cellular body fluids was given during the period of antidiuresis. While the patient was on the comparatively high mineral diet, convulsions could not be induced by the water-retention method, although the values for serum calcium, inorganic phosphorus, and proteins were not significantly different from what they were during the low-mineral dietary period (Figure 3, e). Administration of sufficient parathyroid gland extract during the period of low mineral intake to produce a fairly marked degree of hypercalcemia for a few days prior to and during the period of post-pituitary antidiuresis protected the subject completely from convulsions, although the total amount of water retained in this instance exceeded that in the foregoing experiments (Figure 3, d). The intravenous administration of killed typhoid bacilli at intervals, with the production of symptoms simulating those of an acute febrile illness of bacterial origin, resulted in the occurrence of generalized convulsions during the period of low mineral intake but had no such effect during the period of adequate mineral consumption (Figure 3, b and f). Fever induced artificially by means of diathermy failed to provoke convulsions, even when the mineral content of the diet was extremely low (Figure 3, c). In this instance a large negative water balance resulted from excessive sweating as indicated by changes in body weight. The only significant alteration found in the blood constituents was a slight increase in serum calcium at the end of the febrile period. Ingestion by the subject of sufficient extra sodium bicarbonate during the period of high mineral intake to give the diet a decidedly alkaline ash resulted in no apparent increase in his convulsive reactivity even when typhoid vaccine or pitressin was administered (Figure 3, g and h).

The differences in the patient's response to the various experimental procedures used in connection with the two dietary regimens were obviously related to the difference in total mineral intake. Examination of the data on mineral balances during the different periods, as illustrated in Figure 2, shows that calcium, magnesium, and phosphorus were in negative balance during the period of higher intake. During the latter period

## IRVINE McQUARRIE

calcium and magnesium retentions were relatively higher than that of phosphorus. Since the values for the important serum constituents showed no significant variations attributable to the dissimilarity in diet during the first two experimental periods, one is practically forced to the conclusion that the difference in the convulsive tendency was due to different effects of the diets on the tissues of the brain. Smith and his coworkers (17) demonstrated that prolonged use of a low-mineral diet, similar to that employed in the present study, caused an increase in the water content of various organs. While data for brain tissue were not presented in their report, it is highly probable that the brain would respond like other organs to this type of dietary deficiency. Calcium salts were found to have a protective, dehydrating effect. It has been reported by Ellis (18) that the water content of the brain is definitely increased in experimental parathyreoprivic tetany. Baar (19) has reported a similar increase in brain tissue water in infantile tetany.

Much of the extra water occurring in the brain under these conditions may be stored within the brain cells. Procedures such as limiting the mineral intake or forcing retention of water without equivalent retention of electrolytes, as in certain of the foregoing experiments, would naturally accentuate the abnormal storage of water due to parathyroid insufficiency. Any associated tendency toward a disturbance in the electrolyte balance on the two sides of the brain cell membranes would likewise be increased. Procedures resulting in relative dehydration of the tissues, on the other hand, such as the use of a high-mineral diet or the production of excessive sweating, would be expected to have an ameliorating effect on the same basis. The beneficial effect of parathyroid gland extract in active tetany would appear to depend upon its dehydrating action, which may be related to the increase in total serum calcium, as well as upon its more specific functions of regulating calcium-ion concentration and phosphorus excretion.

*Factors involved in the mechanism of insulin convulsions.*—While it may be taken for granted that hypoglycemia plays a

## THE MECHANISMS OF CONVULSIVE REACTIVITY

primary role in the causation of insulin convulsions analagous to that of hypocalcemia in parathyreoprivic convulsions, it is apparent from a number of considerations that this factor alone does not comprise the physicochemical mechanism of the convulsive episode per se. The well-recognized variability in the degree of hypoglycemia essential for the induction of convulsions in different experimental animals and in human subjects, together with the observation that coma or even death may occasionally occur at extremely low blood sugar levels without the appearance of a convulsion, indicates that additional factors are involved.

As a result of their studies on the mechanism of convulsions in insulin hypoglycemia, Drabkin and Shilkret (20) and more recently Drabkin and Ravdin (21) concluded that the state of hydration of the animal determines whether or not convulsions will occur. In normally hydrated animals the course of events following the intravenous injection of 20 units of insulin per kilogram of body weight was conceived by these authors to be as follows: hypoglycemia  $\rightarrow$  anhydremia  $\rightarrow$  rise in cerebrospinal fluid pressure  $\rightarrow$  convulsions. Animals "dehydrated" by being deprived of water but not of food for periods of from five to seven days did not have convulsions in spite of severe insulin hypoglycemia. Drabkin and Ravdin reported that excision of the stellate ganglia prevented convulsions, even in well-hydrated animals. The significance of these reports has subsequently been brought into question, however, by Corwin (22), who could not prevent insulin convulsions by depriving dogs of water, and by Phillips and Barker (23), who found both dogs and cats to have convulsions following administration of insulin after excision of the stellate ganglia. The latter authors attribute the results of Drabkin and Ravdin primarily to the fact that they observed their animals for periods of only three to four hours after insulin administration instead of for longer periods. The few experiments which we have carried out on the influence of state of hydration, while not conclusive, have left us with the distinct impression that severe dehydration tends to

## IRVINE McQUARRIE

raise the convulsive threshold somewhat, thereby delaying the onset of convulsions or preventing their occurrence entirely in exceptional instances.

Because of certain similarities in the effects produced on the central nervous system by anoxemia and hypoglycemia (24) the most popular current theory regarding the mechanism of insulin convulsions relates the phenomenon to cerebral anoxia. Glickman and Gelhorn (25) reported that insulin convulsions were more readily precipitated in rats when they were subjected to reduced atmospheric pressure and concluded from this that anoxemia sensitizes animals to insulin hypoglycemia. Comments on these reports and the anoxic theory of hypoglycemic convulsions will be made later in the present report.

The increasing evidence that potassium ions play an important role in nervous activity (26) suggested the possibility that the well-known decrease of this ion in the blood plasma following administration of insulin (27) might be shown to indicate its involvement in the mechanism of insulin convulsions. Accordingly a fairly extensive experimental investigation (28) was undertaken to explore this possibility, the immediate objective being to determine the relationship between the level of serum potassium and the occurrence of convulsions. Incidentally changes in the inorganic phosphorus, lactic acid, glucose, hemoglobin, and other constituents of the blood were likewise determined in some instances.

The effects of insulin alone in doses of 10 to 25 units per kilogram of body weight were determined 93 times in 52 well-standardized dogs. In addition eight experiments were carried out with administration of potassium salts (chloride and citrate) in amounts sufficiently large to maintain the serum potassium at normal levels while the blood sugar and inorganic phosphorus were reduced by the injection of insulin. In other experiments phosphate as well as potassium was administered to prevent decrease in both of these inorganic ions during the period of insulin hypoglycemia.

The results of these experiments may be summarized very

## THE MECHANISMS OF CONVULSIVE REACTIVITY

briefly as follows: No constant relationship was found to exist between the level of either the potassium or the inorganic phosphorus of the serum and the occurrence of convulsions. There was no tendency for convulsions to occur at the lowest point on the curves for these two constituents that was comparable with that observed in the case of glucose. Maintenance of the serum potassium and phosphorus at normal levels by the administration of salts of these elements during periods of insulin hypoglycemia did not prevent convulsions. Nor did the extra salts administered increase the convulsive tendency detectably.

On the basis of the anoxic theory of insulin convulsions the convulsive reactivity should be greatly increased by a marked reduction in the amount of  $O_2$  going to the brain, and the relative significance of changes in the various blood constituents might be expected to become more apparent under such circumstances. With this possibility in mind a fairly comprehensive study was made on the effects of breathing atmospheres varying in their partial pressures of  $O_2$ ,  $N_2$ , and  $CO_2$  (29). The method employed was that of determining the effects of insulin on the different blood constituents and on the occurrence of convulsions first in room air and subsequently when the experimental animal was placed in an ordinary oxygen tent at atmospheric pressure with different gaseous mixtures.

The results of these studies may be summarized briefly as follows: It was found, quite unexpectedly, that atmospheres with low  $O_2$  content (4.5 to 12.5 per cent  $O_2$  plus 95.5 to 87.5 per cent  $N_2$ , instead of increasing the convulsive reactivity after insulin administration, tended either to prevent seizures entirely or to delay their onset significantly in a great majority of the experiments. The lower the  $O_2$  pressure within the range noted above, the more effective was the procedure in preventing convulsions. However, when animals were subjected to atmospheres of pure or nearly pure  $N_2$ , generalized convulsions occurred within two or three minutes, even when no insulin was given.

The decrease in convulsive tendency resulting from the ani-

## IRVINE McQUARRIE

mal's breathing an atmosphere low in  $O_2$  was apparently not due to changes involving the other blood constituents, because the potassium, inorganic phosphorus, and glucose were all found on occasion to fall to levels well below those recorded during the control experiment but without the occurrence of convulsions.

As for the effect of breathing an atmosphere high in  $CO_2$  (15 per cent  $CO_2$  plus 20 per cent  $O_2$  plus 65 per cent  $N_2$ ) on the response to insulin administration, it was found that convulsions could frequently be prevented entirely by this procedure so long as the  $O_2$  content was not above 20 per cent. Replacement of most of the  $N_2$  in this mixture by  $O_2$ , however, was found to nullify the anticonvulsive effect of the high  $CO_2$  atmosphere very largely.

Representative experiments are shown in Figure 4. From the upper panel of the figure it will be seen that dog 30 had a convulsion during the control experiment approximately 5 hours after the insulin was administered and when the blood glucose was about 25 mg. per 100 cc. In contrast with this result he did not have a convulsion within 10 hours when he was placed in a low  $O_2$  atmosphere 3 hours after receiving the same dose of insulin. It will be observed that the blood glucose fell to 15 mg. per 100 cc. and that the potassium and inorganic phosphorus of the serum were reduced to levels well below those recorded during the control experiment. The effects of breathing 15 per cent  $CO_2$  are shown in the lower panel of the figure. It will be seen that the results of breathing an atmosphere consisting of 15 per cent  $CO_2$ , 20 per cent  $O_2$ , and 65 per cent  $N_2$  in general resemble those of the low  $O_2$  experiment. Although the blood sugar was ultimately reduced during the experiment to a level well below that at which a convulsion had occurred in room air, no sign of a seizure was observed. The only notable difference in the responses of the blood constituents in the two types of experiment shown in Figure 5 was the prompt and striking elevation of serum inorganic phosphorus resulting from inhalation of  $CO_2$ . That the central nervous system was not

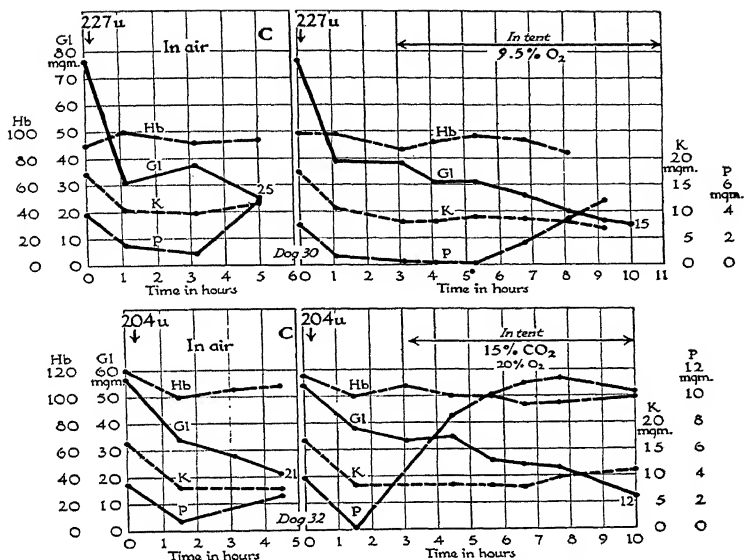


Figure 4. Hyperinsulinism (normal dogs). Effects of breathing atmospheres with low partial pressure of O<sub>2</sub> or with high partial pressure of CO<sub>2</sub> on blood constituents and on occurrence of convulsions after administration of insulin. (Drawn in the Medical Art Shop, University of Minnesota.)

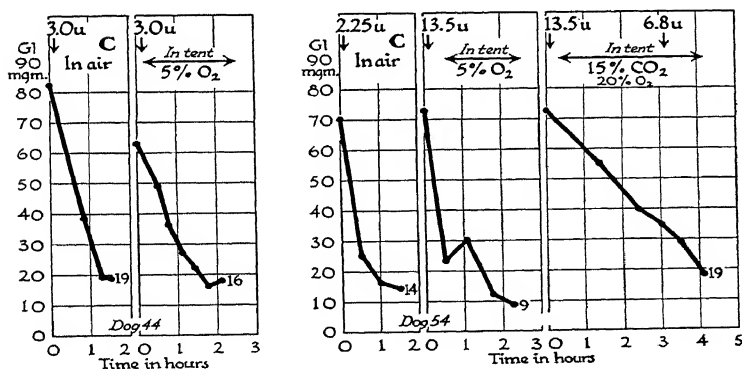


Figure 5. Hyperinsulinism (adrenalectomized dogs). Effectiveness of low O<sub>2</sub> and of high CO<sub>2</sub> atmospheres in preventing convulsions in spite of marked hypoglycemia. (Drawn in the Medical Art Shop, University of Minnesota.)

## IRVINE McQUARRIE

rendered entirely refractory by the prolonged inhalation by the animal of the low  $O_2$  or the high  $CO_2$  atmosphere was indicated by the fact that typical convulsions could be induced within two or three minutes by the administration of metrazol at any time during the course of the experiment.

Because adrenalectomy is known to increase an animal's sensitivity to insulin between 25 and 50 times, a number of the foregoing experiments were repeated in bilaterally adrenalectomized dogs with the cooperation of Dr. M.R. Ziegler, Dr. W.E. Stone, Dr. O.H. Wangenstein, and Dr. Clarence Dennis (30). The results of these experiments may be summarized briefly as follows: Convulsions were induced after adrenalectomy by as little as 1 per cent of the large dosage of insulin used before operation. Whereas a period of between 4 and 5 hours was usually required to induce convulsions in the normal animal, even when an extremely large dose of insulin was given, they occurred within 1 or  $1\frac{1}{2}$  hours after insulin in the adrenalectomized animal. Under the latter conditions the blood glucose was reduced much more rapidly than either potassium or inorganic phosphorus. The latter though slightly reduced were often still within the range of normal when convulsions occurred, indicating the greater significance of hypoglycemia.

Breathing an atmosphere with an extremely low  $O_2$  content resulted in the development of a fairly marked degree of hyperglycemia in the normal fasting animal, whereas it regularly caused hypoglycemia after bilateral adrenalectomy. In spite of this fall in blood sugar no convulsions occurred in adrenalectomized animals subjected to the low  $O_2$  atmosphere. This was true even when the degree of hypoglycemia, so produced in a given dog, was as great as or greater than that observed in the same animal following a convulsive dose of insulin before the adrenals were removed. Breathing an atmosphere consisting of 15 per cent  $CO_2$ , 20 per cent  $O_2$ , and 65 per cent  $N_2$  caused a mild degree of hyperglycemia instead of hypoglycemia in the adrenalectomized dog (see Figure 5).

The foregoing results furnish no support for our original con-



## THE MECHANISMS OF CONVULSIVE REACTIVITY

jecture that some relationship might exist between changes in the plasma electrolytes, particularly potassium and phosphates, and the occurrence of insulin convulsions. Of the various blood constituents determined, glucose alone showed a constant relationship to convulsions when they occurred.

The most significant and surprising result of this investigation was that pertaining to the inhibitory effect of reduced  $O_2$  tension on convulsive reactivity due to overdosage of insulin. In this respect the effect of breathing 15 per cent  $CO_2$  was practically identical with that of the low  $O_2$  atmosphere so long as the partial pressure of  $O_2$  was maintained at a value below that in ordinary air. The fact that convulsions occurred in spite of the 15 per cent  $CO_2$ , whenever the  $O_2$  content of the atmosphere was extremely high, suggests that the inhibiting effect of the  $CO_2$  may depend upon the same general mechanism as that responsible for the low  $O_2$  effect. Further investigation will be required to elucidate the true nature of the phenomenon.

*General discussion.*—The above data, pertaining to the convulsive mechanism in genuine epilepsy, point to the probability that some inherent defect in the brain cells of persons suffering from this disorder is responsible for the spontaneous development of abnormal electrical potentials in the brain, which result in the occurrence of seizures at irregular intervals. From the relatively meager and indirect evidence at present available it appears likely that the convulsive phenomena are dependent upon a profound disturbance in the surface functions of the cortical cells. This tentative interpretation is based upon the abnormal response of the epileptic subject to forced dilution of his extracellular body fluids and upon the circumstance that there appears to be a direct relationship between convulsive reactivity and permeability of brain cell membranes to electrolytes. Studies on the blood lipids in epilepsy have furnished additional indirect evidence in support of the hypothesis that the brain cell membranes are characteristically defective in this disorder (31).

The contention that  $Ca^{++}$  concentration in the extracellular

body fluids is not the only factor involved in the convulsive mechanism of parathyreoprivic tetany is confirmed by the experimental data summarized in the foregoing report. The results of several of the special experiments presented are consistent with the view that a disturbance in the water balance of the brain plays a significant role in this connection.

In the light of the striking parallelism between the convulsive response of a subject with *active* tetany and that of one with genuine epilepsy to such factors as forced dilution of the extracellular body fluids and induced alkalosis, the idea that the type of cerebral dysfunction involved in the convulsive reaction may be essentially the same in the two conditions would appear to justify further experimental investigation. On the basis of this hypothetical interpretation the primary disturbance in the metabolism of calcium, phosphorus, and water which is associated with *active* hypoparathyroidism would appear to produce some temporary alteration in the cells of the central nervous system closely resembling the inherent defect postulated as characterizing the brain cells of the epileptic subject. As demonstrated by the present study, prolonged subsistence on a dietary regimen characterized by its extremely low mineral allowance tends to accentuate the existing abnormality, whatever its nature may be, whereas persistence in the use of a diet high in certain of the essential minerals serves to ameliorate it.

As for the convulsive mechanism in hyperinsulinism, it appears obvious from the data presented above that neither hypoglycemia nor anoxia per se accounts satisfactorily for the occurrence of convulsions. Attempts to correlate changes in the potassium, inorganic phosphate, and other constituents of the blood with convulsive reactivity have thrown no positive light on the question. However, since insulin convulsions could not be induced without hypoglycemia, it appears highly probable that the latter state results in a temporary alteration in the brain cells analagous to that postulated as occurring in *active* tetany and in idiopathic epilepsy. Breathing atmos-

## THE MECHANISMS OF CONVULSIVE REACTIVITY

pheres of low  $O_2$  content or of high  $CO_2$  content tends to prevent this change. The observation of Davis (32) and of Hoagland (33) that the electrical activity of the cerebral cortex is reduced as a result of anoxemia may have an important bearing upon this phase of the subject. The more recent work of Gelhorn confirms the inhibiting effect of hypoxemia on the convulsive mechanism (34).

Results of recent experimental studies by Yannet (35) on cerebral changes in prolonged hypoglycemia are very pertinent to the present discussion. Briefly, it was found from his investigations that the brains of cats examined from three to five hours after injection of large doses of insulin showed a significant decrease in the sodium and chloride content with no change in total water, nonlipoid nitrogen, and potassium. This finding was interpreted as signifying a shift of water from the extracellular to the intracellular space, leading to cellular hydration. The brains of animals examined twenty-four hours after prolonged severe hypoglycemia, which produced histological changes in the ganglion cells (condensation and degeneration), showed an increase in the chloride and sodium content, a marked decrease in potassium, and no significant changes in the total water and nitrogen. These results were interpreted by their author as evidence of cellular shrinking with loss of water and potassium and probably a gain in sodium. It was his impression that these latter changes were those to be expected in tissue undergoing dissolution.

The foregoing sketchy discussions taken together suggest that the convulsive mechanisms involved in the various disorders studied are all based upon similar disturbances in the surface functions of the brain cell membranes. At the present time the author and his collaborators are preparing to test this general hypothesis more thoroughly by determining, so far as possible, movements of "tagged" or radioactive isotopes of potassium, sodium, and phosphorus across the brain cell membranes under appropriate experimental conditions.

*Summary and conclusions.*—1. Results of recent experi-

## IRVINE McQUARRIE

ments on the convulsive mechanisms in parathyreoprivic tetany and hyperinsulinism are described, and previously reported data pertaining to genuine epilepsy are reviewed.

2. The fact that convulsions can be induced in the epileptic but not in the normal subject by procedures which result in dilution of the extracellular body fluids is interpreted tentatively as indicating the existence of an inherent defect in the brain cell membranes in this disorder.

3. Other evidence presented likewise suggests that the characteristic abnormality in epilepsy involves a disturbance in the surface functions of the brain cell membranes.

4. This supposed membrane dysfunction is postulated as being responsible for the spontaneously occurring disturbances in the rhythm of electrical potentials in the brain, which characterize this chronic convulsive state.

5. As regards hypoparathyroidism, further evidence is furnished in support of the contention that other factors, in addition to decreased calcium-ion concentration, play significant roles in the causation of convulsions in this condition.

6. The convulsive reactivity in parathyreoprivic tetany was shown to be markedly increased by the prolonged use of a low-mineral diet and significantly decreased by a high mineral intake.

7. During the period of low mineral intake the parathyreoprivic subject was found to respond to forced dilution of his extracellular body fluids and to other epileptogenic procedures in a manner almost identical with that of patients suffering from genuine epilepsy.

8. As for the mechanism of insulin convulsions, evidence is presented which indicates that no direct relationship exists between the decrease in serum potassium or inorganic phosphorus and the occurrence of convulsions following administration of insulin in large doses. Of the various blood changes recorded, hypoglycemia alone was found to bear a direct relationship to convulsions.

9. No support for the anoxic theory of insulin convulsions

## THE MECHANISMS OF CONVULSIVE REACTIVITY

was found. On the contrary, experimental data are presented which indicate clearly that hypoxemia within certain limits greatly diminishes the convulsive reactivity. It was found that convulsions either did not occur at all or were markedly delayed in their onset if an animal was kept in an atmosphere with greatly decreased  $O_2$  content or with high partial pressure of  $CO_2$  after the insulin was injected. This was found to be true even when the blood sugar was reduced well below the predetermined convulsive level for a given animal.

10. Breathing low  $O_2$  or high  $CO_2$  atmospheres was found to prevent insulin convulsions in bilaterally adrenalectomized dogs, even when the dose of insulin was from four to eight times the predetermined convulsive dose for such insulin-sensitive animals in ordinary room air.

11. From these data the deduction is drawn that some other factor in addition to hypoglycemia, possibly a specific type of alteration in the brain cell membrane resulting from hypoglycemia, is responsible for the convulsion. It appears that partial deprivation of  $O_2$  interferes with the operation of this unknown factor.

12. It is tentatively concluded that the convulsive mechanisms in idiopathic epilepsy, parathyreoprivic tetany, and hyperinsulinism are all explainable in part on the basis of disturbances in the surface functions of the brain cell membranes.

### *References*

1. (a) Berger, H. 1932. Ueber das Elektrenkephalogramm des Menschen. *Arch. f. Psychiat.* 97:6. (b) Fischer, M.H. 1933. Elektrobiologische Auswirkung von Krampfgiften am zentralen Nervensystem. *Med. Klin.* 29:15. (c) Kornmüller, A.E. 1935. Der Mechanismus des epileptischen Anfalles auf Grund bioelektrischer Untersuchungen am Zentralnervensystem. *Fortschr. d. Neurol. Psychiat.* 7:391, 414. (d) Gibbs, F.A., Davis, H., and Lennox, W.G. 1935. The electro-encephalogram in epilepsy and in conditions of impaired consciousness. *Arch. Neurol. Psychiat.* 34:133. (e) Gibbs, F.A., Gibbs, E.L., and Lennox, W.G. 1937. Epilepsy: a paroxysmal cerebral dysrhythmia. *Brain* 60:377.
2. Lennox, W.G., and Cobb, S. 1928. Epilepsy from the standpoint of physiology and treatment. *Medicine* 7:105.

# IRVINE McQUARRIE

3. Lennox, W. G., and Gibbs, F. A. 1936. Oxygen saturation of the arterial blood in epilepsy. *Arch. Neurol. Psychiat.* 35:1198.
4. Gibbs, F. A., Lennox, W. G., and Gibbs, E. L. May, 1939 (in press). Carbon dioxide and epilepsy. *Proc. Am. Psychiat. Assn.*
5. Teglbjaerg, H. P. Stubbe. 1932. Epilepsy and hypoxemia. *Acta psychiat. et neurol.* 7:659.
6. Cobb, S. 1932. Causes of epilepsy. *Arch. Neurol. Psychiat.* 27:1245.
7. Georgi, F. 1926. Zur Pathogenese des epileptischen Anfalls. *Zeitschr. f. Neurol.* 106:751.
8. Syz, H. C. 1927. Observations on experimental convulsions with special reference to permeability changes. *Am. Jour. Psychiat.* 7:209.
9. Spiegel, E. 1931. Central mechanism of generalized epileptic fits. *Am. Jour. Psychiat.* 10:595.
10. McQuarrie, I., Manchester, R. C., and Husted, C. 1932. Study of the water and mineral balances in epileptic children. *Am. Jour. Dis. Child.* 43:1519.
11. Spiegel, E. A., and Spiegel-Adolf, M. 1939. Physiochemical mechanisms in convulsive reactivity. *Jour. Nerv. Ment. Dis.* 90:188.
12. McQuarrie, I., and Peeler, D. B. 1931. Effects of sustained pituitary antidiuresis and forced water drinking in epileptic children. *Jour. Clin. Invest.* 10:915.
13. Ziskind, E., and Somerfeld-Ziskind, E. 1939. Hydration studies in epilepsy. *Jour. Nerv. Ment. Dis.* 89:52.
14. Dragstedt, L. R. 1927. The physiology of the parathyroid glands. *Physiol. Rev.* 7:499.
15. Mathien, F. 1937. Recherches sur l'insuffisance parathyroïdienne expérimentale. *Arch. internat. de physiol.* 44:516, 529.
16. McLean, F. C., Barnes, B. O., and Hastings, A. B. 1935. Relation of parathyroid hormone to the state of calcium in the blood. *Am. Jour. Physiol.* 113:141.
17. Eppright, E. S., and Smith, A. H. 1938. Changes in water of tissues induced by diets containing various mineral supplements. *Am. Jour. Physiol.* 121:379.
18. Ellis, M. M. 1928. Water content of certain tissues during acute guanidine and parathyroprivia tetanias. *Biochem. Jour.* 22:937.
19. Baar, H. 1928. Bertrag zur Kenntnis der Beziehungen zwischen Tetanie und Wasserhaushalt. *Zeitschr. Kinderheilk.* 46:502.
20. Drabkin, D. L., and Shilkret, H. 1927. Insulin anhydremia; importance of water-reserve in physiological crisis. *Am. Jour. Physiol.* 83:141.
21. Drabkin, D. L., and Ravdin, I. S. 1937. The mechanism of convulsions in insulin hypoglycemia. *Am. Jour. Physiol.* 118:174.
22. Corwin, W. C. Unpublished data, personal communication.
23. Phillips, R. A., and Barker, S. B. 1938. Insulin convulsions after removal of the stellate ganglia. *Am. Jour. Physiol.* 124:202.
24. (a) Omsted, J. M. D., and Logan, H. D. 1923. The effect of insulin on the central nervous system and its relation to the pituitary body. *Am. Jour. Physiol.* 66:437. (b) Kleitman, N., and Magnus, R. 1924. Zur

## THE MECHANISMS OF CONVULSIVE REACTIVITY

- Wirkung des Insulins auf das Zentralnervensystem. Arch. f. d. ges. Physiol. 205:148. (c) Omsted, J. M. D., and Taylor, A. C. 1924. The effect of insulin on the blood. Am. Jour. Physiol. 69:142. (d) Himwich, H. E., Bowman, K. M., Wortis, J., and Fazekas, J. F. 1937. Brain metabolism during the hypoglycemic treatment of schizophrenia. Science 86:271.
25. Glickman, N., and Gelhorn, E. 1938. The effect of oxygen deficiency on the sensitivity of rats to insulin. Am. Jour. Physiol. 121:358.
  26. Holmes, E. G. 1935. The metabolism of brain and nerve. Ann. Rev. Biochem., Stanford Univ. Press 4:435.
  27. (a) Briggs, A. P., and Koechig, I. 1923. Some changes in the composition of blood due to the injection of insulin. Jour. Biol. Chem. 58:721. (b) Harrop, G. A., Jr., and Benedict, E. M. 1924. The participation of inorganic substances in carbohydrate metabolism. Jour. Biol. Chem. 59:683.
  28. Ziegler, Mildred R., and McQuarrie, Irvine. 1938. Mechanism of insulin convulsions. I. Serum electrolytes and blood sugar in relation to occurrence of convulsions. Proc. Soc. Exper. Biol. Med. 39:523.
  29. McQuarrie, Irvine, and Ziegler, Mildred R. 1938. Mechanism of insulin convulsions. II. Effects of varying partial pressures of atmospheric  $O_2$ ,  $N_2$  and  $CO_2$ . Proc. Soc. Exper. Biol. Med. 39:525.
  30. McQuarrie, I., Ziegler, M. R., Stone, W. E., Wangenstein, O. H., and Dennis, C. 1939. Mechanism of insulin convulsions. III. Effects of varying partial pressures of atmospheric gases after adrenalectomy. Proc. Soc. Exper. Biol. Med. 42:411.
  31. (a) McQuarrie, I., Bloor, W. R., Husted, C., and Patterson, H. A. 1933. Lipids of blood plasma in epilepsy. I. Statistical study of single determinations in one hundred epileptic and thirty-two "normal" subjects. Jour. Clin. Invest. 12:247. (b) McQuarrie, I., Husted, C., and Bloor, W. R. 1933. Lipids of blood plasma in epilepsy. II. Variations in relation to occurrence of seizures. Jour. Clin. Invest. 12:255. (c) Cowie, D. M., and Magee, M. C. 1935. Lipoid content of tissues in cases of status epilepticus, toxic encephalopathy and chronic leptomeningitis. Am. Jour. Dis. Child. 49:884. (d) Aird, R. B., and Gurchot, C. 1939. Protective effect of cholesterol in experimental epilepsy. Arch. Neurol. Psychiat. 42:491.
  32. Davis, P. A., Davis, H., and Thompson, J. W. 1938. Progressive changes in the human electroencephalogram under low oxygen tension. Am. Jour. Physiol. 123:51.
  33. Hoagland, H. 1938. Brain metabolism and brain wave frequencies. Am. Jour. Physiol. 123:102.
  34. Gelhorn, E., and Packer, A. C. 1939. Influence of anoxia on glycogenolytic action of adrenalin. Proc. Soc. Exper. Biol. Med. 41:345.
  35. Yannet, H. 1939. Experimental study of cerebral changes in prolonged hypoglycemia. Am. Jour. Dis. Child. 58:918.

# METHODS OF ANALYSIS OF NERVOUS ACTION

BY

Herbert S. Gasser

*Director, the Rockefeller Institute for  
Medical Research*

IN THE seventeenth century Glisson defined irritability. Glisson was an experimenter and a philosopher. Glisson the experimenter showed that a contracting muscle does not increase in size, and thus once and for all dispelled the notion that contraction is occasioned by an inflow of animal spirits. Glisson the philosopher pondered on the activity of organs. He observed that the gall bladder discharges bile when it is irritated. Then he argued that if organs can be irritated, they must have the power of being irritated. And he called this indwelling power *irritability*.

There is no small measure of tautology in this argument. No hint is given about what actually happens. No procedure is suggested. As the statement stands it has about the same practical value as the idea, long current, that the manifestations of the brain and its connections are attributable to an indwelling something called mind.

Many years passed, during which Glisson's statements remained unnoticed. Then they were brought forth from the oblivion into which they had fallen and hailed as a milestone in the history of physiological thought. In the meantime irritability had been rediscovered by Haller, and the meaning of the term had been extended by later physiologists to refer to any change which takes place in a tissue under the influence of an exciting agent.

To a large extent, however, I fear that irritability has been but a term to which to cling because of the intellectual comfort it affords. A tissue reacts because it is irritable. This in effect



## ANALYSIS OF NERVOUS ACTION

has been said in countless lectures, and the subject has been dismissed with a sigh of relief and even an air of satisfaction. In so far as this has happened the term has been an anesthetic rather than what it should be, an irritant.

The merit in the idea lies in the fact that it is the first step toward the analysis of nervous action. There is supplied a convenient reference word, the significance of which transcends its content of meaning. It marks out a void that must be filled.

The history of neurophysiology is characterized by much astute reasoning, stimulated by the obscurity with which the nature of the nerve impulse has been surrounded and prompted by the desire to penetrate the obscurity to something intellectually tangible. Illuminating ideas have appeared long in advance of the time at which they could be proved or disproved. Many of these ideas have been embodied in the form of models. For these models physiologists are often criticized—justly and unjustly. Justly, because models can belie their proper function as tools in empirical research and defeat the methods of science by becoming vehicles for the projection of a creation of the observer's mind upon reality. Unjustly, because models in their place are valuable. When a biological phenomenon can be imitated in a physical or chemical system, our faith is strengthened that a physicochemical interpretation of the phenomenon can be found. Observations are made and fragments of information are gained. Integrated into the model these fragments acquire a wider meaning, and the model, enriched, acquires a greater usefulness.

Nerve models have now been developed to a degree of complexity which prevents their construction outside of the imagination. It is necessary to combine the core-conductor principle of Hermann with a concentration cell of Nernst or a sieve membrane of Ostwald, and then in some way to assimilate the combination with the iron-wire model of Lillie, in order to arrive at even a skeleton representation of the membrane hypothesis. The modern era of nerve physiology started with the membrane hypothesis. And so well has the hypothesis served that at no

## HERBERT S. GASSER

time during its history has it found a competitor. Today it stands more firmly entrenched than ever before.

As you know, according to the tenets of the hypothesis, a nerve fiber is considered to be a fluid conductor, surrounded by a labile surface film maintained in a steady state through the expenditure of energy derived from oxidation in the fiber. Across the film a potential gradient is maintained by an ionic-concentration difference at its borders, and possibly by other factors. During activity a molecular change is supposed to occur in the surface, with a resultant fall of the potential gradient, a flow of current from the adjacent segments, and excitation of these segments by the current—the whole process being repeated throughout the length of the fiber and constituting the nerve impulse. Such in brief is the theoretical background for the physical and chemical measurements that make up the analysis proper.

Haller, if he had been confronted with a piece of nerve not connected with an effector organ, could not have told by any test that he could make whether the fibers were active or inactive. At that time there was no possibility of measuring to the least degree changes in potential, in heat production, or in chemical constitution. Nor was it possible for a long time to come. Scientific curiosity drove physiologists to make experiments, but no success could meet their attempts when, as in the years following Galvani, they tried to measure tissue potentials with a gold-leaf electroscope. Oersted's discovery, Ampère's laws, and the Schweigger multiplier had to intervene before Nobili could build a galvanometer sensitive enough to be moved by a bioelectric current.

Galvanometers, devices for controlling the electric current, and the developments in the growing science of electrochemistry determined the course taken by electrophysiology in the nineteenth century. The outcome was the laws of excitation, an understanding of the mechanism of polarization, and a knowledge of the steady and the slowly changing potentials in tissues. Our libraries contain the records of much beautiful work on the

## ANALYSIS OF NERVOUS ACTION

spread of polarization and the distribution of potential about a generating focus—work to which little attention is now being paid because of our preoccupation with the action potential. In time a recrudescence of interest in these subjects may be anticipated and then the libraries will yield up their treasures, unless, as may happen, science follows the wasteful course of rediscovery.

Had it not been for technical limitations nineteenth-century electrophysiology would have gone much further and its history been quite different. Physiologists had an intense desire to determine the course of the action potential, and they exercised great ingenuity in their experiments. Bernstein's rheotome gave them their first answer, but the instrument was too cumbersome to come into general use as a tool of research. Then came the capillary electrometer of Lippmann and the string galvanometer of Einthoven. The experience of many of us goes back to the use of these instruments. We recall their lack of sensitivity and the records that revealed the period of the instrument rather than the form of the impressed potential. At the same time we recall what was accomplished by the men who used them, and we hail these men for the geniuses that they were. They were working against odds. Again a cycle in physical discovery had to run before the odds could be removed. Edison, J. J. Thomson, Richardson, and De Forest—then the vacuum-tube amplifier and the vacuum-tube oscillograph.

At the present time amplifier-oscillograph assemblies are available which are at once fast enough to record without distortion the most rapid potential changes that the nervous system can generate and steady enough to report faithfully the long-continued, small potential changes that follow a period of activity. Furthermore it is possible with differential input stages in the amplifiers to obtain simultaneous records from a number of positions in the same animal without appreciable mutual interference between the leads. In other words, recording has become thoroughly satisfactory for all potentials \* that can be

\* In order to avoid criticism of the use of the term *potential* it must be explained

detected. Detection, however, is subject to limitations, and there is no indication that they can be removed. The limitation is set by a series of random spontaneous potentials collectively known as the *noise level*. As given by the engineers, the root-mean-square value of the noise level is a little over a microvolt. From a practical standpoint, however, it is larger than this figure. Six microvolts would be a better general estimate. The faster the amplifier the greater the apparent disturbance, which simply means that the disturbance is recorded instead of being hidden. It would often be convenient to be able to record potentials smaller than the noise level, but as this is impossible, means of compensating for the difficulty must be resorted to. Fortunately, the local potential differences at the point where they arise are not small. The small sizes at the positions of the recording electrodes are attributable to the interposition of inactive tissue acting as a shunt. Compensating devices, therefore, depend upon avoiding shunts; in nerve, cutting down the size in order to record potentials in single small fibers, as in the Adrian-Bronk technique; in the central nervous system, bringing needle electrodes into accurate apposition with the active nuclei or fibers. When advantage is taken of practices like these the natural limitation set upon the size of a detectable potential change is not a serious handicap.

When the technique for the measurement of any one sign of activity is much easier to apply than that of any other sign, there is always a tendency to use that technique to the exclusion of the others and, therefore, to develop a one-sided view of the mechanism being studied. The very fact that we use the word *electrophysiology* implies that we may be bringing forth a gospel of the nervous system according to the electrode. For that reason we should pause long enough to inquire into the status of the questions that we are asking and of the answers that we are obtaining with respect to the needs of plain physiology without any modifier.

that for the sake of convenience it has become customary among physiologists to use the single word in the sense of the potential difference between two points.

## ANALYSIS OF NERVOUS ACTION

In the first place, a potential change reveals that something is happening. Other events showing the same thing are changes in excitability, metabolism, heat production, and electrical impedance. Measurements of excitability have been in use for the longest time. Action is accompanied by changes in the threshold of excitation. But unless the nerve is connected to an effector organ, the thresholds cannot be measured without the use of one of the other signs of activity. Of these, the electrical sign is the only one having practical importance.

Correlated to the demonstration of processes of activity is the ability of the potential technique to measure the intensity and form of these processes when modified by various controlled conditions, such as anesthesia, states of oxidation, and the effects of pharmacodynamic agents or of alteration of the ionic composition of the medium.

Of all the questions that may be asked, those best answered by the potential technique relate to time: When? How long? The answers to these questions are very precise and supply the principal data for the analysis of many problems. The number-frequency interpretation of the mechanism of the nerve message depends upon them, likewise the knowledge of how impulses pass synapses, and of how central ganglionic masses recover from activity. Without these answers we would not know the sequence of events in the activity cycle. And we would know very little about the fiber composition of nerve trunks. Many other instances could also be cited.

Questions about time are not answered with precision in experiments undertaken for the measurement of any of the other signs of activity. Among the latter, the best localized sign is that of heat production. In experiments of great technical perfection A. V. Hill and his colleagues have been able to determine the course of the heat production with an accuracy sufficient to show that more than 95 per cent of the heat is liberated after the message-carrying part of the activity is over. But there is still doubt about what happens during the spike process. Even in a cooled frog nerve the spike lasts only a few milliseconds,

and the heat analyses are not able to work down to such short periods.

Measurements of metabolic changes usually require destruction of the tissue, and the question of time is not raised. The exceptions are in relation to gaseous metabolism. In the experiments that have been performed to determine the absorption of oxygen and the production of carbon dioxide in isolated nerve, the metabolism has been brought only into a very rough time relationship with the period of activity. A better localization has been accomplished, however, in the mammalian cerebral cortex by Dusser de Barenne, McCulloch, and Nims, who used the glass electrode to measure hydrogen-ion concentration changes—changes which of course depend upon the diffusion of carbon dioxide between the tissue and the electrode. Here it was possible to correlate the concentration changes with slow potential changes. For measurement of the rate of oxygen utilization in muscle G. Millikan has devised a differential photoelectric colorimeter method; and Urban and Peugnet have recorded instantaneous changes in the absorption spectrum with the aid of the electron oscillograph. It is possible that after elaboration and refinement these techniques may be made applicable to nerve, inasmuch as cytochrome is present in nerve fibers. If the question of the time of production were raised in connection with the nongaseous metabolites, in all probability some sort of answer could be gained through a more extended employment of the technique of rapid immersion into liquid air, as it is used by von Muralt and others.

Not so good an answer is obtained to the question: Where? Interest in this question has grown during recent years, as electrophysiologists have transferred their attention from isolated nerves to the central nervous system. As long as they were working with isolated nerves, localization was not much of a problem, as they were dealing essentially with a linear system surrounded by an insulating medium. In the central nervous system, on the other hand, they are dealing with systems of unknown shapes

## ANALYSIS OF NERVOUS ACTION

imbedded in an extensive conducting medium. This fact makes the problem much more complex. No difficulty, it may be said, is experienced with respect to the degree of localization that is necessary for following excitation through the pathways of the nervous system. The difficulty begins when localization has to be precise enough to permit using the form of the recorded potential in the interpretation of what is taking place.

A word of explanation must be introduced at this point in order to make clear why the form of the action potential is of interest. The details of the form are known through the parts of the nervous system that are readily accessible to observation — the nerve fibers. Three kinds of nerve fibers exist, which differ somewhat in their properties; but in all three the configuration of the action potential is in accord with a basic pattern. The action starts out with a large, short negative deviation known as the spike, and is continued with a relatively prolonged set of changes of low voltage known as the after-potentials. In their simplest form the after-potentials start out with a negative component (in one fiber class this component is vestigial) and are continued by a longer positive component. During the period of the negative component the fibers are more excitable than when they are at rest, and during the period of the positive component they are less excitable.

Numerous bits of evidence support the view that in the central nervous system the action potential of the neurons follows the basic pattern, to the extent at least of being made up of a spike followed by a positive after-potential. As a corollary to this view it would follow that during the period of the positive after-potential in a postganglionic fiber it would be more difficult for impulses to pass the synapse. In favored positions, where the relationship can be tested, this correlation has been found to hold.

It is just here, however, that the difficulty begins to enter. Factors other than subnormality of the postganglionic neurons can prevent transmission across a synapse. And it is often al-

## HERBERT S. GASSER

most impossible in actual records to identify the parts of the potential cycle because of the uncertainty about where the potential is generated.

The neurons in the gray matter are so short with respect to the wave length of any of the electrical events that for practical purposes all parts of the neuron may be considered to be in phase. From this fact it follows that for any potential change to be recorded at all, there must be an intrinsic differential between the parts of the neuron with respect to the magnitude of the change. Activity in a neuron pool, therefore, would be characterized by the setting up of a closely interlocked system of sources and sinks oriented in numerous directions. Positivity in a record may mean a real positive change in the neurons under the exploring electrode, or appropriately oriented negative changes in adjacent neurons. Whether it is the one or the other makes the greatest possible difference in the interpretation of the records.

The way out of the difficulty seems to be along two lines. First, a better understanding of potential theory. Physiologists have now begun to pick up the threads laid down by Helmholtz and Hermann more than half a century ago. Secondly, it is doubtful that, even armed with a full understanding of potential theory, it would be possible to arrive at a unique solution for the positions of the sources and sinks. In addition it will be necessary to have much better information about the details of cellular organization of the gray matter than we now possess.

There remains now one final question: What? What are the processes that take place? To this question electrical recording does not give an answer. It is well always to keep clearly in mind that one cannot infer a process from a potential. The potential is a guide, and interpretations depend upon what other events can be correlated with it. Future progress depends upon the perfection of other techniques, more rapid methods of measuring heat, and, above all, finer and more rapid methods for determining chemical changes. It is impossible, for me at least, to predict the directions from which advances may be



## ANALYSIS OF NERVOUS ACTION

expected, but a number of techniques that are yielding valuable information at the present time may be mentioned.

Among the instructive experiments of recent years are those of Cole and Curtis, in which the impedance changes in the nerve fibers were measured. They show that during action there is a large change in the resistance of the plasma membrane, while, contrary to the view previously held, the change in capacity is small. Furthermore, they demonstrate with great clearness the time in the action potential at which the changes in the surface of the fiber take place. The recorded potential, up to the point of inflection in the rising phase of the spike, is simply an expression of the polarizing effect of the bioelectric current upon the segment of nerve ahead of the active region. During this period, in other words, conduction resembles that in a submarine cable. It is only when the polarization reaches a certain point that the active change takes place, revealing itself through a fall in the transmembrane resistance and, of course, through an addition to the negativity of the potential obtaining in the region.

The experiments of Cole and Curtis were performed upon single giant axons of the squid. These fibers, which were introduced into the equipment of physiologists by J. Z. Young, have also been proved otherwise highly favorable objects for study, along with single axons of Crustacea. The material is particularly valuable for study of the processes that take place at the locus of a stimulus before the spike emerges and begins to be conducted (Hodgkin), and for measurements of the resistance of the plasma membrane (Cole and Hodgkin). Schmitt and Baer have been able to draw off the axoplasm and to examine it chemically. And in a recent personal communication Dr. Hodgkin informed me that he has been able to introduce a micro-pipette electrode into the inside of the fibers and to record the action potential directly across the plasma membrane.

The other methods deal with structure rather than with action. Their objective is to extend the resolution of form beyond the range that is possible with the ordinary microscope. The

## HERBERT S. GASSER

techniques employed involve the use of polarized light, X-ray diffraction, and the electron microscope. The progress that has been made with the aid of the first two techniques has recently been reviewed by Schmitt and Baer. Not a great deal has as yet been learned about the axon, but it has been possible to formulate a provisional picture of the ultrastructure of the sheath, in which the sheath is seen to be made up of concentric layers of protein laid down at intervals of 171 Å and separated by two bimolecular layers of lipid. What the electron microscope will be able to contribute remains to be determined.

Thinking about nervous action always brings one promptly to a consideration of events taking place in and about very thin films. Therefore much is to be expected from the developments that are taking place in two-dimensional chemistry and from the work upon built-up films of the Langmuir type.

It is scarcely any longer prophetic to look forward to a biology in molecular dimensions, for there is no small amount of contemporary interest in intramolecular and intermolecular patterns, as witness the recent history of the myosin fibril. There is a growing body of knowledge relating to the subject, based upon direct and indirect information derived from physical and chemical methods.

In nerve physiology the goal toward which we are turned is a complete understanding of nerve mechanics. It is not too soon to demand knowledge of the position of the molecules in the fibers, nor is it too audacious to inquire into how molecules move during activity. Of all the highly differentiated and highly active structures of the body a nerve fiber is one of the simplest. Hence, if any biological structure is to be proved not too complex as a physical system to be subjected to successful analysis, it should be this one. At any rate, until answers have been obtained to the questions that have been posed, full meaning content cannot be put into the term *irritability*.

# THE NERVOUS REGULATION OF VISCERAL PROCESSES

BY

Detlev W. Bronk

*Director, Eldridge Reeves Johnson Research  
Foundation and the Institute of Neurology,  
University of Pennsylvania*

MAN'S life is dominated by his environment. Every change in the physical surroundings initiates a sequence of events within his body, and thus his behavior is modified from moment to moment. How wide a range of environmental stimuli the human organism can survive depends upon its power to maintain the normal course of vital processes despite the changed external conditions. The limits of human endeavor are accordingly determined by the capacity of the body to adjust its activity to varied surroundings.

In order to widen the scope of his life man endeavors to control his environment. He mitigates the rigors of the seasons by regulating the temperature of his surroundings. By artificial illumination he increases the duration of his active day. He ascends to high altitudes with the aid of artificially prepared oxygen.

It is customary to speak of such achievements as "man's conquest of nature." The phrase is an arrogant misrepresentation. Stated more truly, the scientists have formulated certain laws describing natural phenomena, and they have thereby enabled us to utilize the forces of nature for our needs and our desires. Actually, this knowledge makes possible a richer life only when it is used to bring the human organism into a more satisfactory relationship with its external surroundings. A humanitarian society will therefore seek to direct its technological developments so as to satisfy the biological requirements of the individual.

There is a growing appreciation of this need for creating a

more hospitable environment for man. It is a concern that has indeed been forced upon society by the machine age, for a machine is a device whereby man alters his relations to his environment and thus changes the conditions of his life. Gradually there has come a realization that we have been giving more thought to the machine than to the biological consequences of the machine-made environment. Threatened domination by our technological creations is at last forcing us to consider the biological consequences of travel at fast speeds and high altitudes, to study the effects on the organism of work in gaseous atmospheres unnatural to man, to investigate the physiological results of new routines of life required by our industrial system. Far-sighted engineers recognize biological science as an important part of their training. Architects are beginning to think not only of the materials they employ but also of the biological needs of the living organisms who are to inhabit their structures.

These are signs of great promise for human welfare. They represent a movement toward the "biotechnic civilization" predicted by Patrick Geddes and Lewis Mumford (1). But if this biological accent in technology is in fact to improve the welfare of mankind, we must acquire and disseminate a better understanding of how the physical environment modifies the activity of the body. This is a primary social responsibility of the biologist and physician.

In the simplest forms of organic life all of the vital activities are under the direct domination of external conditions. In the more highly developed organisms, however, most of the cells are not under the immediate influence of the external environment. The cells live amid relatively stable surroundings and are accordingly able to perform their specialized functions in a precise manner. "The stability of the *milieu intérieur* is the primary condition for their freedom and independence of existence" (2). This freedom is, however, only relative, for the stability of the internal environment depends upon the continual adjustment of the activity of each individual cell to the varying

## NERVOUS REGULATION OF THE VISCERA

conditions external to the organism. The behavior of every cell of the body is governed by the physical and chemical characteristics of the outside world, but the control is exercised primarily through the agency of the sense organs, through the integrative action of the central nervous system, and finally through the effects of the motor nerve cells. Thus a harmony of action is achieved within the assemblage of units which preserves for each of them an environment that is favorable for the cells' specific functions.

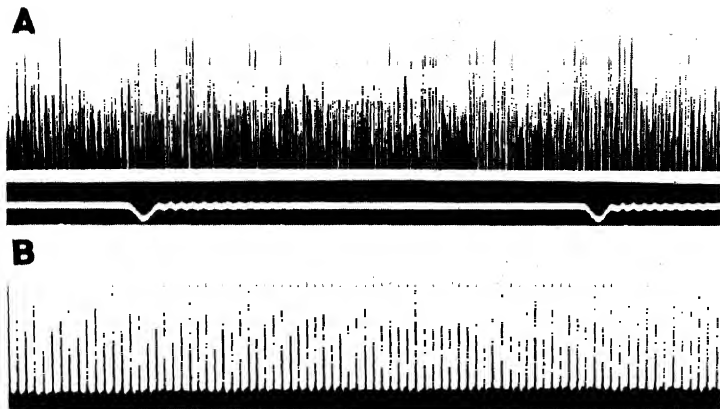
The relations of the nervous system to the structures which it controls have been revealed by the microscopic resolution of the body into its constituent cells. Only thus has it been possible to gain a clear picture of the anatomical basis for the nervous mechanisms discovered by the physiologist. To understand the functional characteristics of those mechanisms it is necessary to resolve the activity of the nervous system into the events which take place within the individual neuron. How a sense organ or a motor nerve cell is modified by a change in its environment, how it signals to its distant termination a report of the disturbance, how, there, it activates another cell—these are fundamental questions concerning the nervous control of the organism.

The "trigger" for this sequence of events in a nerve cell is some adequate variation in the physical or chemical surroundings. In this way the cycle of chemical reactions within the neuron is modified, and the unstable molecular structure is altered. The changed properties of this disturbed section then act as an altered environment for the adjoining regions, and their structure is accordingly modified. When finally this propagated wave of physical and chemical change reaches the terminations of the neuron the modified properties of the terminations alter the environment and thus the activity of the adjoining muscle or gland or nerve cell. This is the quantum of nerve action which drives the cells of the organism. I need not discuss at greater length the characteristics of this impulse for that has been the subject of the preceding paper by Dr. Gasser. As he has shown,

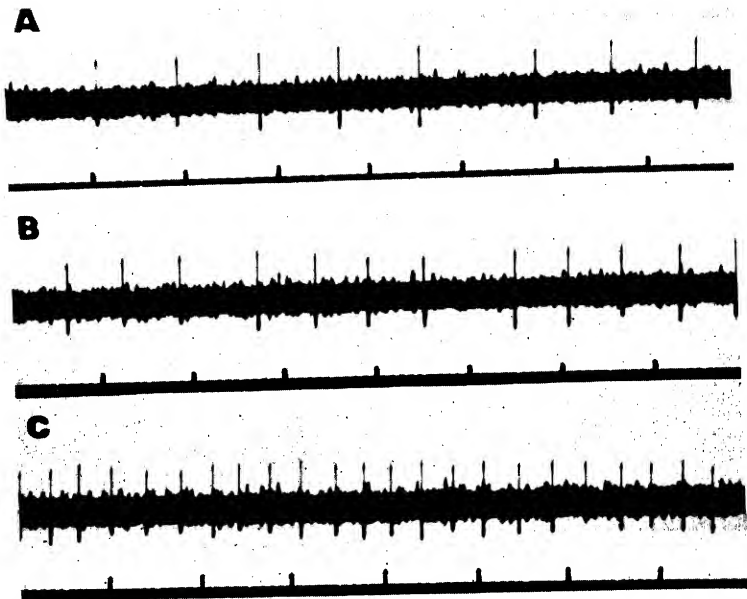
the impulse is most satisfactorily initiated for experimental study by passing a brief electric current through the nerve which is under observation.

In the living organism, however, nerve cells are exposed to fluctuations of their environment that are usually less transient. This is especially true of the stimulating action of chemical agents. The general character of the response of a neuron to such a change in its surroundings can be shown by bathing a short stretch of a nerve fiber in a solution of sodium citrate, which removes calcium from the molecular structure of the nerve. As a result of this continuously acting chemical stimulus there develops not a steady change of state throughout the fiber but a succession of impulses sweeping along at constant intervals, and the electrical signs of these impulses can be recorded in the manner described by Dr. Gasser. The persistent alteration of structure at the site of calcium removal acts as a continuing stimulus for the adjoining portion of the fiber. However, that adjoining region of the fiber, and each succeeding region, reacts not continuously but with a cycle of change in which reorganization must follow breakdown before the persistent stimulus can again act. This rhythmic sequence of events is the characteristic response of a unit of the nervous system to a change in its environment. To be observed it is necessary to record the activity of a single neuron, for the dissimilar responses of a group of nerve cells obscure the character of their individual behavior. This contrast is illustrated in Figure 1.

It is by such trains of periodically recurring events that the nervous system regulates visceral function. Over a fiber of the vagus nerve, for instance, there goes a succession of rhythmic impulses which produce an inhibitory action when they reach the heart. Concurrently, this inhibition is opposed by the acceleratory effects of trains of impulses arriving over the fibers of the sympathetic system. The muscle cells in the walls of the blood vessels and those concerned with respiration are likewise under the control of these periodic neural events. As each im-



*Figure 1. A: Electrical signs of the nerve impulses conducted along a group of nerve fibers from a localized region which has been stimulated by the application of sodium citrate. B: A similar record of the electrical changes in one of the many fibers involved in A. Each vertical line is a photographic record of the movement of the oscillograph caused by a spike potential associated with one nerve impulse. In succeeding figures a spike potential will be referred to merely as "an impulse." Time record: 1 sec.*



*Figure 2.* Impulses discharged from a nerve cell in a ganglion (cat) perfused with acetylcholine in three different concentrations. A: 25 micrograms per cc. B: 50 micrograms per cc. C: 100 micrograms per cc. Frequency of cellular action varies with the concentration of the exciting agent. Time: 1 sec. (Larabee and Bronk, 3.)



## NERVOUS REGULATION OF THE VISCERA

pulse reaches the termination of its nerve fiber it briefly modifies the properties of the visceral cells innervated by the fiber.

From this it must follow that the degree of activity of an organ at any instant is determined by the number of nerve impulses it receives. That in turn depends upon the number of neurons conducting impulses to the organ at the time, and upon the frequency with which the several neurons discharge impulses to their dependent cells. We should expect to find that these two variables in the nerve messages are governed by the state of the external and internal environment, if the nerve cells are to be effective intermediaries between the environment and the organs which must adjust to environmental changes. That this is so can be established experimentally.

Consider, for instance, the response of a motor nerve cell to varying concentrations of a specific chemical agent. For this purpose a sympathetic ganglion has many advantages.

These ganglia have, too, several important roles in the nervous regulation of the viscera, for the impulses that leave the centers by way of the autonomic nervous system must exert their influence through the medium of these structures. Each fiber coming to a ganglion makes contact with many ganglion cells, and thus the visceral influence of a cell within the autonomic centers is made more extensive. Each ganglion cell, on the other hand, is innervated by many preganglionic fibers. There is accordingly a pooling of the effects of the impulses from the cells of the central nervous system as they converge upon these relay stations. The several trains of impulses thus lose their identity in the development of a state of excitation within the ganglion. The number of ganglion cells that are brought into action in this way and the rate at which they discharge their impulses to the viscera are determined by the characteristics of the cells as well as by the number of impulses that come from the centers.

This digression concerning the functions of a sympathetic ganglion may have raised the question how the preganglionic impulses excite the ganglion cells. And that leads us back to the

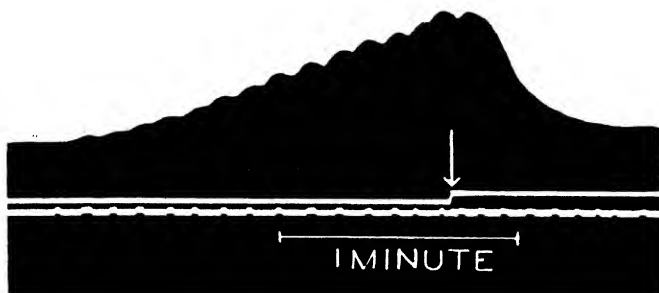
nature of the response of a motor nerve cell to a chemical agent. For, while I must not trespass on Dr. Cannon's subject, I shall divulge no secret when I remind the reader that there are many who believe this excitation in a ganglion to be the result of acetylcholine liberated by the presynaptic impulses.

It will accordingly be of especial interest to employ acetylcholine as the chemical agent with which to stimulate a nerve cell in a sympathetic ganglion. At any one concentration, maintained constant by steady perfusion, the cell under observation discharges over its postganglionic fiber a fairly regular and persistent succession of impulses at a characteristic frequency. As the concentration of acetylcholine was raised from 25 micrograms to 100 micrograms per cubic centimeter of perfusion fluid, in the experiment illustrated in Figure 2, there was an increase in the frequency at which impulses were transmitted out over the axon. Similar effects of other chemicals, such as potassium and calcium-binding sodium citrate, can be readily demonstrated on motor nerve cells and peripheral axons. From such evidence and other that will be cited later we know that the frequency of impulses initiated in a motor nerve cell is a measure of the intensity of the environmental stimulus. By variations in the rate of their activity the individual neurons exercise their regulatory functions.

Beneath the histological similarity of one nerve cell to another of the same type there are certain differences of basic structure that determine the nature of the cells' reactions to their environment. In the several neurons of a group subjected to a certain intensity of stimulus there is accordingly no one characteristic frequency of action. Each unit has its own peculiar rhythm of response that is determined by its properties at any instant. Likewise, there is among a population of nerve cells a wide range of thresholds for excitation. This is the basis for the second type of signaling mechanism employed by the nervous system. As the intensity of a stimulus increases the less excitable cells are stimulated, and thus the number of active cells in the viscera is increased.

## NERVOUS REGULATION OF THE VISCERA

The effects that these rhythmic series of impulses produce when they reach their destination are obviously determined by the nature of the cell that is controlled by the nerve fiber. The effects also depend upon the character of the impulses. Inasmuch as the impulse is by definition a reversible change in the basic structural organization of a nerve fiber, the impulse will necessarily differ from neuron to neuron. In Figure 3, for example, there is represented the effects of a series of impulses discharged over sympathetic constrictor fibers to the muscular



*Figure 3.* Response of blood vessel to series of constrictor (sympathetic) and dilator (dorsal root) impulses. Short volleys of impulses applied to sympathetic fibers every 1.5 sec. (indicated by lower signal). The arrow indicates the beginning of dorsal root stimulation at 20 per sec. (indicated by upper signal). (Bozler, 4.)

walls of a blood vessel. Each volley of impulses produced a slow twitch of the muscle, and the record shows these twitches summing into a slow contraction of the arterial wall. The simultaneous arrival of a train of impulses over fibers that come by way of the dorsal roots inhibited the contraction and caused thereby a relaxation of the muscular walls. We must assume that the terminations of these two sets of fibers possess quite different characteristics. Such differences are most clearly illustrated by the liberation of characteristic substances from specific types of fibers, and that is Dr. Cannon's story.

Because the properties of any one fiber vary from time to time the character of its impulses is not constant. One of the

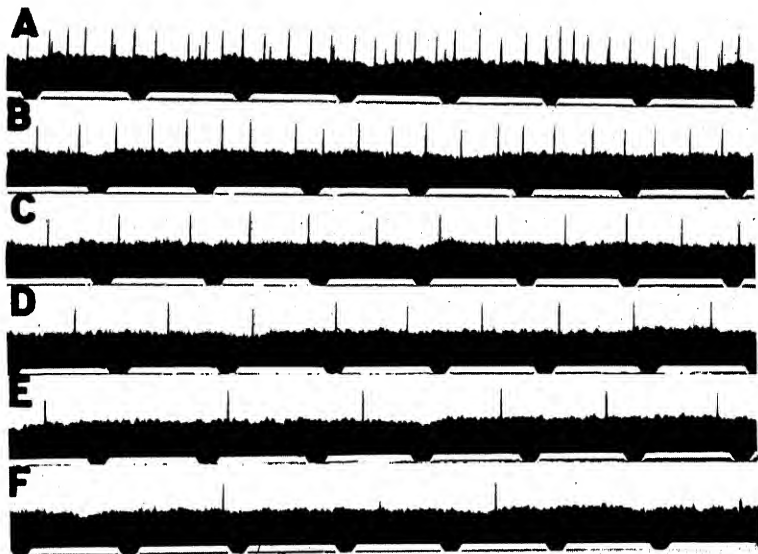
most significant causes of such variations is the previous activity of the neuron, for the nerve must go through a long cyclic process of recovery before it regains its original condition. A familiar manifestation of this is the reduction in size of the electrical sign, or spike potential, of a nerve impulse that follows in the relatively refractory phase of a previous impulse. On the other hand, the spike potential is considerably increased if it is propagated over a nerve fiber during the delayed recovery phase that is identified by the positive after-potential (Gasser, 5). There are reasons for believing that the magnitude of this spike potential is a measure of the effectiveness with which an impulse reaching the termination of a nerve fiber acts upon an adjoining cell. The influence exerted by nerve cells on other nerve cells and ultimately upon the organs of the body depends therefore not only on the number of neurons in action and the frequency with which they discharge their impulses; it depends also upon the timing of the impulses relative to antecedent events in the nerve pathway.

A nerve cell is capable of long-sustained periods of activity at a constant rate, but such behavior is not characteristic of the neurons which control the organs of the body. Trains of impulses are continually starting and stopping, and the frequency of their rhythm varies from moment to moment. There is, for instance, the periodically intermittent outflow of impulses to the respiratory muscles (see Figure 4). And in the sympathetic nerve trunks we observe waves of activity which are due to the periodic waxing and waning of the discharge from large groups of nerve cells in the centers (7, 8). It is this inconstant activity of the visceral nervous system that makes possible the maintenance of a stable environment within an organism subjected to external changes.

To know the factors that govern this variable activity of the motoneurons of the visceral nervous system is to understand the primary influences to which the viscera are subject. They are, in the first place, governed by the internal environment that they stabilize. This activation of the visceral motoneu-



*Figure 4.* The rhythmic discharge of impulses in a single motor nerve fiber to an external intercostal muscle (cat) during two respiratory cycles. Time marker:  $\frac{1}{6}$  sec. Lower line: pneumogram, upward movement indicating inspiration. (Bronk and Ferguson, 6.)



*Figure 5.* Impulses discharged from a motor nerve cell supplying an internal intercostal muscle during asphyxial stimulation. Cat curarized. Respiration pump started at A after prolonged asphyxia. Records at 10 sec. intervals. Time marker:  $\frac{1}{6}$  sec. (Bronk and Ferguson, 6.)

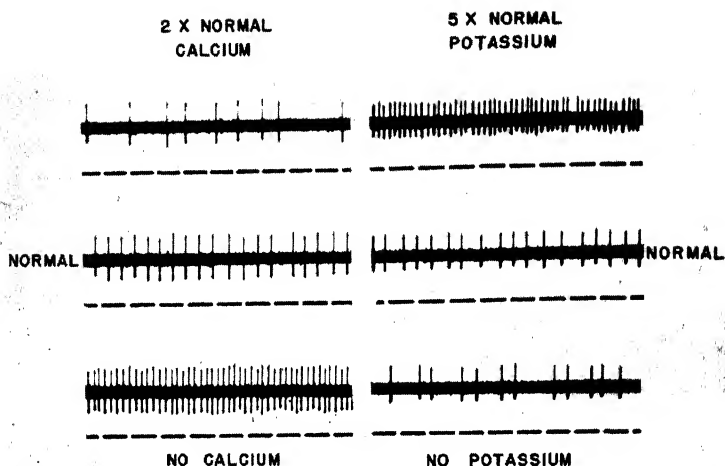


Figure 6. Effects of calcium and potassium ions on the response of a nerve cell (in sympathetic ganglion of a cat) to acetylcholine. Ganglion perfused with solutions containing 40 micrograms of acetylcholine per cc. Middle records are with acetylcholine in normal Ringer's fluid; others modified as indicated. (Larrabee and Bronk, 8.)

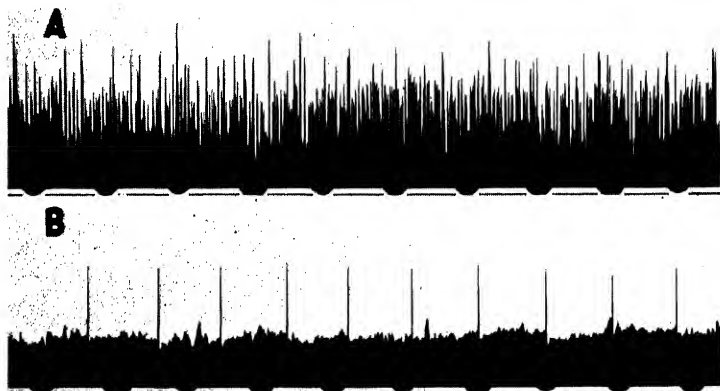


Figure 7. Typical records of impulses discharged from sense organs. A: Impulses in many fibers from a large group of sense organs. B: In one of those fibers from a single receptor. Time:  $\frac{1}{5}$  sec. (Bronk, 7.)

## NERVOUS REGULATION OF THE VISCERA

rons by shifts in the composition of the blood is not surprising in view of the extremely sensitive reaction of most nerve cells to changes in their chemical environment.

Consider, for instance, the effects of asphyxia upon the discharge of impulses from a nerve cell supplying a respiratory muscle. In the experiment illustrated in Figure 5 the respiratory center was isolated as far as possible from nervous influences arising outside the central nervous system, and so I believe we were dealing with the response of a nerve cell to changing properties of the blood. As the asphyxia was prolonged the tempo of events in the neuron was accelerated, owing to the progressively greater modification of the cellular properties by the asphyxial blood. The fibers of the respiratory muscle supplied by this neuron would accordingly be stimulated more frequently, with a corresponding increase in the depth of respiration. The chemical conditions responsible for the increased activity of the controlling neuron are thereby corrected. A similar increase in the rate of impulse discharge during asphyxia has been recorded from the sympathetic motor nerve cells.

There is relatively little of such information concerning the influence of variations in the composition of the body fluids on the activity of the cells in the central nervous system. It is, however, an important problem, for the properties of the nerve cells are continually shifting with even minor and obscure chemical changes in the body. Certain of these are adequate to excite or to inhibit the neurons, and they thus produce grossly manifest changes in the effector organs. But other important chemical influences merely modify the capacity of a nerve cell for excitation by another agent and therefore play a role that is less readily observed. Typical of this action are the effects of changes in the concentration of calcium and potassium ions on the response of a nerve cell to a specific chemical substance such as acetylcholine (see Figure 6). The degree of response to a certain concentration of this substance is profoundly influenced by the relative concentrations of those two ions. Likewise, the ability of a cell to respond to impulses in

other nerve cells across a synapse depends upon the nature of the chemical environment at any instant. Any consideration of the nervous regulation of visceral action necessarily requires an appreciation of the continually shifting influence of the fluids bathing the cells of the nervous system.

A more specific and localized control of the motoneurons is provided by the sense organs. These outposts of the nervous system are characterized by an extraordinary sensitivity to certain changes in their surroundings. Those sense organs which are densely scattered over the surface of the body provide a relationship between the motor nerve cells and the physical forces that threaten the constancy of the internal environment. Those located in the walls of the blood vessels, in the roots of the lungs, and throughout the interior of the body report to the control centers of the nervous system the state of the organs and those variations of internal environment that demand correction.

How these reportorial functions are performed can best be determined by intercepting and recording the electrical signs of the messages transmitted over the afferent fibers. The record obtained from a bundle of sensory nerves is a confusion of independent messages because no two sense organs respond in precisely similar manners to a certain stimulus. To decipher the record it is therefore again necessary to deal with a single neuron (see Figure 7). When we do so, we find that a sense organ which is under the influence of an unvarying stimulus dispatches a series of rhythmically recurring impulses. The signaling mechanism employed by the sensory neurons is accordingly the same as that of the motor neurons. Indeed, it is difficult to see how it could be different inasmuch as the conducting pathways have essentially the same characteristics.

The similarity continues when we go on to consider what variable in the sensory message connotes a change of stimulus intensity. How, for instance, do the sense organs in the carotid sinus — an important reference station for the determination of blood pressure — report the arterial pressure to the cardiovas-



## NERVOUS REGULATION OF THE VISCERA

cular regulatory centers? To answer this question we have recorded in an afferent nerve fiber the impulses discharged from one of the receptors while perfusing the sinus with blood maintained at a constant pressure (see Figure 9). When this procedure is repeated at various pressures we find that the frequency of rhythmic activity in the nerve is a measure of the pressure. The frequency turns out to be an important variable in the sensory nerve message.

It is, however, not a direct measure of environmental conditions, for a sensory neuron has no one characteristic frequency of response to a given intensity of stimulus. The rhythmic activity of a receptor under a constant and persistent stimulus



*Figure 8.* Gradual decrease in the rate of impulse discharge from a single sense organ (photoreceptor) under constant intensity of stimulation which continues during blacking of signal line. Time signal:  $\frac{1}{2}$  sec.

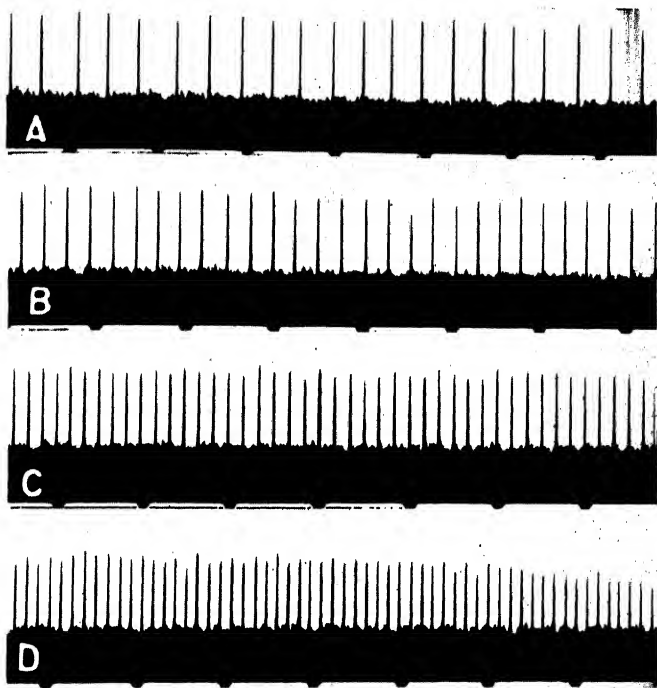
does not continue at a constant rate. On the contrary, there is a progressive decline in the frequency with which impulses are discharged (see Figure 8). In a manner characteristic of most living cells, a sense organ is modified by a change of the environment in such a way that it responds less actively as time goes on to a constant environmental stimulus. In some receptors this process of adaptation is so rapid that only a few impulses are evoked at the beginning of a stimulus. Such are the tactile receptors which are useful only for the signaling that is associated with movements. The visceral receptors whose characteristics are known are, however, of the slowly adapting variety. This is fortunate, for otherwise they would not give adequate reports of the slow variations in the state of the vascular and respiratory systems.

But even here there is no constant relation between the intensity of stimulus and the frequency of impulses in the mes-

sage transmitted by a sensory nerve fiber. And so the visceral nervous system has no absolute measure of steady conditions. This is, however, not a serious defect in the nervous regulatory mechanisms, for static conditions do not generally exist within the body. The primary function of the visceral receptors is to report changes in the state of the organism, and for that task they are quite adequate.

If all sense organs were equally sensitive to changes in their surroundings, a change of impulse frequency would be their only means for varying the activity of the central nervous system in accordance with the conditions they find throughout the organism. But there are marked differences of individual excitability—owing presumably to differences in the basic structure of the receptors. As the level of arterial pressure or the distention of the lungs increases, the thresholds of the less irritable receptors are reached. The activity of the nervous system is accordingly graded by variations in the number of sensory pathways conducting impulses from the periphery, as well as by fluctuations in the rate at which the rhythmic events occur within the individual neurons.

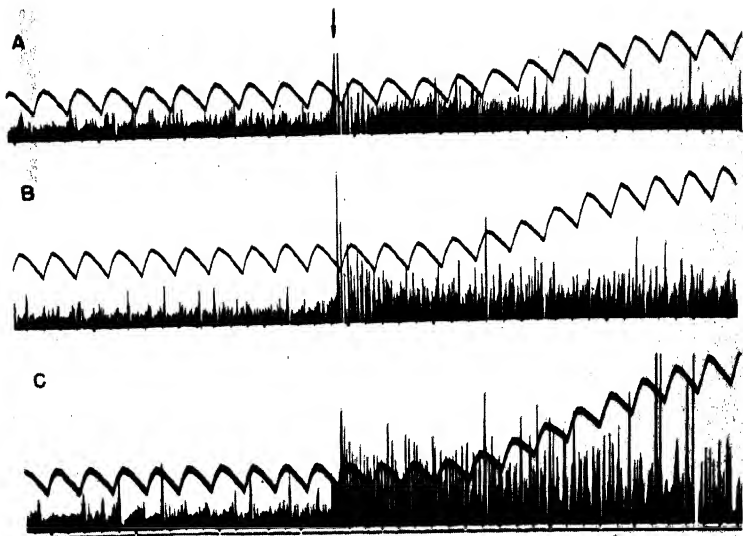
These differences in the sensitivity of receptors also make it possible for the nervous system to perform specific functions. We have seen that the rhythmic activity of sense organs in the walls of the carotid sinus increases with the level of blood pressure. There are also such receptors in other strategically important locations—the aorta and mesentery. Because the properties of nervous tissue are as a rule altered by chemical agents, and especially by changes in the concentration of carbon dioxide and oxygen, we might expect to find the activity of these sense organs modified by the level of those substances. It is noteworthy that this is not the case. For, if they did so, the sensory messages relating to blood pressure would be falsified by alterations in the oxygen tension of the blood. It is characteristic of the effective precision of the mechanisms of visceral control that there are two different groups of receptors in the region of the carotid sinus and at the base of the heart (12).



*Figure 9.* Afferent impulses discharged from a single end organ in the carotid sinus (cat) stimulated by constant pressures within the sinus. In A: 40 mm. Hg; B: 80 mm. Hg; C: 140 mm. Hg; D: 200 mm. Hg. Time marker gives  $\frac{1}{6}$  sec. (Bronk and Stella, 8.)



*Figure 10.* Variations in the frequency of afferent impulses from a pressure receptor in the carotid sinus signaling variations in the arterial pulse pressure. (Bronk and Stella, 11.)



*Figure 12.* Increased discharge of sympathetic impulses in the inferior cardiac nerve (cat) and rise of blood pressure produced by stimulating hypothalamus with a frequency of 150 per sec. Beginning of stimulation indicated by arrow. Relative strengths of stimuli: A, 1; B, 2; C, 3. Time:  $\frac{1}{2}$  sec. (Bronk, Pitts, and Larrabee, 15.)



*Figure 13.* Discharge of impulses in a single fiber of the cervical sympathetic nerve trunk caused by stimulation of the hypothalamus at various intensities and at a rate of 100 per sec. Strength of stimulus increased progressively from A to D. Time:  $\frac{1}{2}$  sec. (Bronk, Pitts, and Larrabee, 15.)

## NERVOUS REGULATION OF THE VISCERA

One set is remarkably sensitive to shifts in the chemical composition of the blood, and the impulses which they initiate exert an important control over the nerve cells that innervate the respiratory muscles. The pressure receptors, on the other hand, are little affected by the composition of the blood. They are accordingly able to signal with remarkable fidelity the variations of blood pressure (see Figure 10).

Through the medium of these nerve messages the activity of the visceral motoneurons is adjusted to the requirements of the organism. Inflation of the lungs increases the rate at which impulses are received by the respiratory centers from the pulmonary stretch receptors (13). The activity of the inspiratory motoneurons is thereby inhibited and the expiratory neurons are stimulated. Distention of the carotid sinus or aorta increases the discharge of impulses from the pressure endings in the vessel walls, and thus the outflow of accelerator impulses to the heart and of constrictor impulses to the blood vessels is checked as in Figure 11. Such a relationship between visceral afferent and visceral efferent neurons brings the nervous regulation of the organs under the control of the organs themselves. The viscera are thus able to guard the limits of their own activity while participating in the coordinated functions of the whole organism.



*Figure 11.* Inhibition of the discharge of sympathetic impulses in the inferior cardiac nerve (middle record) during distention of the carotid sinus. Intra-carotid pressure represented by uppermost record. Time:  $\frac{1}{6}$  sec. (Bronk, 14.)

These protective reflexes operate through the integrative action of nerve cells located in the medulla. But the correlation of the activity of every organ with the activity of every other part of the body suggests that the visceral motoneurons also come under the influence of the higher coordinating centers of the brain. We must accordingly include impulses from those

centers in our consideration of the factors that regulate the activity of the visceral nerve cells — as well as chemical agents and the afferent impulses from external and internal sense organs.

The hypothalamus, for instance, is known to play an important part in processes that involve profound changes in visceral activity, such as temperature regulation and emotional reactions. It will therefore be appropriate in this discussion to investigate the influence of the nerve cells in the hypothalamus on the activity of the visceral motoneurons. This can be done by recording the impulses in a sympathetic nerve while stimulating certain regions of the hypothalamus. Figure 12 shows the discharge of impulses from sympathetic motor nerve cells that were being excited by impulses from the hypothalamus. There is also shown in the figure the rise of blood pressure caused by the sympathetic vasoconstrictor impulses.

The nature of this control can be analyzed more precisely by observing the influence of the hypothalamus upon the activity of a single sympathetic nerve cell. The frequency with which impulses are discharged over a sympathetic fiber is thus found to vary with the strength and frequency of hypothalamic stimulation (see Figure 13). The cardiovascular system is accordingly controlled by the hypothalamus through variations in the activity of the sympathetic motor nerve cells. By similar procedures it is possible to trace the relations of the visceral nervous system to various regions of the brain, and thus it is possible to gain a clearer understanding of how the activity of the organs is integrated into the coordinated actions of the body.

Gradually and laboriously the interpretation of the life of an organism in terms of cellular processes yields a more reasonable conception of biological mechanisms. Such a cellular conception of human activity may never provide an adequate comprehension of man's behavior. But a clearer understanding of the factors which shape the life of man will certainly make possible the creation of a physical and social environment more favorable to his well-being.

## NERVOUS REGULATION OF THE VISCERA

### *References*

1. Mumford, Lewis. 1938. *The Culture of Cities*. Harcourt, Brace, New York.
2. Bernard, C. Quoted from Fulton, J. F. 1930. *Selected Readings in Physiology*. Thomas, Springfield, Ill.
3. Larrabee, M. G., and Bronk, D. W. In Bronk, D. W. 1939. Synaptic mechanisms in sympathetic ganglia. *Jour. Neurophysiol.* 2:380.
4. Bozler, E. 1936. The mechanism of the inhibitory action of vasodilator nerves. *Am. Jour. Physiol.* 117:457.
5. Gasser, H. S. 1937. The control of excitation in the nervous system. *Harvey Lectures* 32:169.
6. Bronk, D. W., and Ferguson, L. K. 1935. The nervous control of intercostal respiration. *Am. Jour. Physiol.* 110:700.
7. Adrian, E. D., Bronk, D. W., and Phillips, G. 1932. Discharges in mammalian sympathetic nerves. *Jour. Physiol.* 74:115.
8. Bronk, D. W., Ferguson, L. K., Margaria, R., and Solandt, D. Y. 1936. The activity of the cardiac sympathetic centers. *Am. Jour. Physiol.* 117:237.
9. Bronk, D. W. 1934. The mechanism of sensory end organs. *Proc. Assn. Res. Nerv. Ment. Dis.* 15:60.
10. Bronk, D. W., and Stella, G. 1935. The response to steady pressures of single end organs in the isolated carotid sinus. *Am. Jour. Physiol.* 110:708.
11. ——— 1932. Afferent impulses in the carotid sinus nerve. *Jour. Cell. Comp. Physiol.* 1:113.
12. Schmidt, C. F., and Comroe, J. H., Jr. 1940. Functions of the carotid and aortic bodies. *Physiol. Rev.* 20:115.
13. Adrian, E. D. 1933. Afferent impulses in the vagus and their effect on respiration. *Jour. Physiol.* 79:332.
14. Bronk, D. W. 1934. The nervous mechanism of cardiovascular control. *Harvey Lectures* 29:245.
15. Bronk, D. W., Pitts, R. F., and Larrabee, M. G. 1939. The role of the hypothalamus in cardiovascular regulation. *Proc. Assn. Res. Nerv. Ment. Dis. Vol.* 20.

# THE ARGUMENT FOR CHEMICAL MEDIATION OF NERVE IMPULSES

BY

Walter B. Cannon

*George Higginson Professor of Physiology,  
Harvard University*

WHEN an impulse travels along a nerve it is attended by a quick rise of negative electrical potential, followed by a quick fall. The duration of this *spike potential* includes the absolutely refractory and the relatively refractory periods of nerve function. The fall is not immediately to zero but is checked by a slower, negative after-potential. That in turn is followed by a longer, positive after-potential (see Figure 1). The changes are more rapid in fibers of large diameter than in small fibers; the spike lasts only 0.4 msec. in the fastest fibers (e.g., those supplying skeletal muscles) and thereupon the nerve can be stimulated again (1). As elsewhere in the body, such electrical phenomena are signs of physicochemical or chemical processes accompanying functional activity. In nerve there must be to a large extent a restoration of the resting state, when an impulse has passed, before another impulse can traverse the same course. Associated with the electrical phenomena of nervous activity is a use of oxygen, an output of carbon dioxide, and a display of heat. Since a nerve soon ceases to transmit impulses in the absence of oxygen, it is reasonable to assume that the increased metabolism, demonstrable when a nerve functions, indicates that chemical work is involved. According to evidence furnished by Schmitt and Gasser (2), the negative after-potential depends on 'active oxidation. It is depressed by asphyxia, and after asphyxia it is much augmented by conditions favoring oxidation. And when this potential is augmented the rate of recovery is hastened. It is highly probable, therefore, that even in the subsidence of the spike potential—for



## MEDIATION OF NERVE IMPULSES

in fibers of fast conduction the negative wave "is first seen at its maximal value" (1)—an oxidative restorative process occurs, essential to returning to the nerve the ability to act again.

The spike potential, sweeping along a nerve which has been effectively stimulated, causes an electric current to run, in a conductor external to the nerve, from inactive points to the active region, the *action current*. It is assumed that a polarized

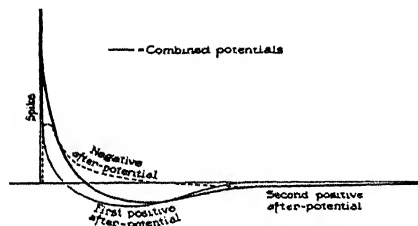


Figure 1. Diagram depicting a theoretical series of potentials which, summed together, would reproduce the action potential. (From Erlanger and Gasser, 1937.)

state exists in the surface membrane of each nerve fiber (the outer charges positive); that a stimulus decreases or abolishes the potential difference across the membrane at a node of Ranvier, making it permeable and thus depolarizing it; and that thereupon adjacent nodes of the fiber discharge through the node already discharged so that new nodes in turn become permeable, depolarized, and activated, and cause next proximal nodes to do the same. That the circulating local currents resulting from this traveling negative electrical wave can cause stimulation of successive portions of a nerve has been proved by Hodgkin (3), who has demonstrated that when an impulse reaches a region blocked by cold or compression, the local currents extend beyond the block and set up there a change of potential with an accompanying change of excitability. Blair and Erlanger (4) have brought supporting evidence by showing that an action current will pass an inert polarized region of the nerve and stimulate the responsive segment beyond. According to the traditional theory of transmission of influence from neurones to other neurones (as in a sympathetic ganglion) or from neurones to the cells of an effector organ (a muscle or

gland), that is the sort of process which occurs: it is supposed that the local circuits reach beyond the limits of the active nerve and excite, electrically, the next element in the series.

At the synapse between neurones and effectors there is evidence of a protoplasmic discontinuity. As Bronk and Brink (5) and also Forbes (6) have pointed out, this introduces a physical condition not found in Hodgkin's experiments. Forbes has remarked, "The structural and presumably the electrical conditions are quite different in an unbroken but inactive axon from those at the synapse. Here, at the termination of the neurone, histology seems to reveal a transverse membrane, which may well act as a short circuit to the action potential. . . . If the membrane theory of nerve conduction holds good, there is every reason to expect such a short-circuiting effect at the termination of the axon." Furthermore, Rosenblueth and I (7, 8) have shown that smooth muscle—that of the nictitating membrane—though readily responsive to a single nerve volley, is, after being deprived of its nerves, quite unresponsive to a single electric shock. Similarly the cells of a gland (the adrenal medulla in our tests) easily activated by impulses delivered by their nerve supply, are, after denervation, affected to only a very slight degree even when electric shocks are applied so powerful and prolonged as to injure the tissue. Obviously there are synapses at which the possibility of stimulation by the action current appears to be very slight.

In recent years evidence has accumulated that when an impulse arrives at a nerve ending it sets free a chemical substance—adrenaline at sympathetic synapses, acetylcholine at parasympathetic synapses and at synapses in motor end plates and in sympathetic ganglia—and that these substances rouse in the next distal element its typical reaction—secretion, contraction, relaxation, or nervous conduction. Two theories of synaptic transmission have been proposed: the traditional electrical theory and the newer chemical theory. Each theory has its ardent advocates, and a vigorous controversy prevails between them. For purposes of the present discussion these advo-

## MEDIATION OF NERVE IMPULSES

cates may be called the electragonists and the chemagonists — agonist meaning contestant or combattant!

There is agreement of the agonists on certain points. In order to simplify the consideration of these points, and also the points of difference, I propose to limit the present discussion to synapses in which there is evidence that acetylcholine is involved. The liberation of this substance at parasympathetic synapses and at the synapses of sympathetic ganglia and motor end plates, on the arrival of nerve impulses, is frankly admitted by the electragonists. They also admit that tissues where acetylcholine appears contain a cholinesterase which quickly destroys that highly unstable combination of choline and acetic acid. They admit further that acetylcholine, when perfused in very minute amounts through organs in which it is discharged by nerve impulses, will influence these organs as do the impulses themselves. They admit likewise that when present in excess acetylcholine has not a stimulating but a paralyzing action. And, finally, they admit that the acetylcholine naturally set free at parasympathetic endings in slowly reacting structures — e.g., in the heart, smooth muscle, and glands — is the agent which affects these structures. The main difference now concerns the role of this mediator in the transmission of influence from nerves to skeletal muscle cells, and from nerves to nerve cells at ganglionic synapses. Electrical stimulation will, indeed, excite both muscle cells and nerve cells; and therefore the action current, if it reaches with effective intensity beyond the junctions, would act as a stimulus. But, if that is the mode of transmission, why the discharge of acetylcholine by the nerve impulse? This is where theories begin to conflict.

In evaluating two antagonistic theories nothing is to be gained by a recital of the phenomena which both can explain. Deeper insight into the processes concerned in the theories is likely to result from a survey of the phenomena which only one, or neither, can account for. On that basis we may proceed to compare the two theories — electrical and chemical — as explanations of observed facts.

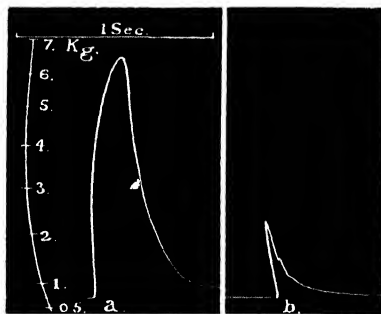
## WALTER B. CANNON

First, the analogy between evidence for chemical mediation in the slowly responsive heart and evidence for it in rapidly responsive striated muscle and sympathetic ganglia. In all three structures a perfusate obtained while they are not stimulated contains no demonstrable acetylcholine. In all three, nerve-stimulation causes the appearance of acetylcholine,



*Figure 2.* Leech muscle. Blood from ganglion with natural circulation: A, C, and F without stimulation; B, D, and E during preganglionic stimulation. (From Feldberg and Vartiainen, 1934.)

*Figure 3.* Spinal cat. Tension record from gastrocnemius, perfused with warm Locke's solution. a: contraction in response to injection by arterial cannula of 20 $\gamma$  acetylcholine. b: maximal motor nerve twitch. (From Brown, Dale, and Feldberg, 1936.)



though eserine may have to be used to prevent its sudden destruction by cholinesterase (see Figure 2). In all three, perfusion with acetylcholine induces the typical effects of nerve impulses — inhibition of the heart beat, contraction of skeletal muscle (see Figure 3), and discharge of ganglion cells. And in all three the effect of acetylcholine can be blocked by drugs, without preventing its liberation at nerve endings — in the heart by atropine, in striated muscle by curare, and in the ganglion by nicotine. Under these conditions, with nerve impulses delivered to the tips of the active nerve fibers, as shown by the discharge of acetylcholine there, and with the recipient elements still sensitive to electrical stimulation, the action current is without influence.

## MEDIATION OF NERVE IMPULSES

The practical identity of evidence for chemical mediation in the three regions raises pertinent questions. If the action current is the true transjunctional stimulus, what is the use of acetylcholine, regularly produced during stimulation? How is the efficacy of minute amounts of acetylcholine, as a stimulus to ganglion cells and muscle fibers, to be accounted for? Why should there be a concentration of cholinesterase in sympathetic ganglia and at the motor end plate (cf. 9, 10)? The electragonists surely face difficulties in answering these questions, difficulties not encountered in the chemical explanation. The orthodox electragonists, moreover, are inconsistent; they agree that acetylcholine is a deputy of nerve impulses at vago-cardiac synapses, but deny it that function for neuromyal synapses. Furthermore, they confront a dilemma: on the one hand, agreement that acetylcholine is a chemical mediator of nerve impulses, and on the other, admission that the widespread arrangements for its routine production and prompt destruction at synapses, where it can stimulate, are merely a false and futile show. In attempted avoidance of this dilemma hypothetical suggestions of imaginary functions have been offered, e.g., that acetylcholine may serve to increase excitability at synapses, or to delay the onset of fatigue, or to dilate blood vessels locally, or to exert vague "trophic influences" (11). Perhaps it "may"; but it can demonstrably do things to muscle cells and to sympathetic neurones; it can excite in them their peculiar functions. Confessedly, the argument implied in emphasizing the presence and capacity for action of acetylcholine at synapses is based on the reasonableness of means' being adapted to ends. In biology, however, that argument is so generally valid as to be, in special cases, quite respectable.

The electragonists have set up what they regard as a serious obstacle to the chemical theory in stressing the brevity of the latent period of the postsynaptic elements before action occurs. That is a criticism which demands attention. In the neurones of the superior cervical ganglion the latent period of the most rapid responses ranges from 2 to 4 msec. (12), and in the fibers

of some skeletal muscles it can be shorter, 1 to 2 msec. (11). If acetylcholine is the transmitter it must flash forth and excite in that brief time; and since a single nerve volley induces a single response the acetylcholine must be quickly rendered ineffective during the refractory phase of the responding elements (which lasts in neurones and muscle fibers about as long as the latent period) before another nerve volley can act again. The time is too short, the electragonists assert, for these chemical changes to occur. It is pertinent to point out that this is an assumption. Too little is known of the speed of chemical processes at synapses to justify categorical limitations. Furthermore the electragonists neglect the fact that the nerve fiber itself, when it transmits an impulse, must recover before it can transmit another. Evidence indicates, as we have noted, that this recovery is an oxidative chemical process; and in fast mammalian fibers it lasts less than 1.0 msec., i. e., it may be much more rapid than the transmission at ganglionic and muscular synapses, which the electragonists assume to be too rapid to be chemical. Until they can offer better evidence for their assumption they have a frail basis for belittling the speed of intimate molecular changes.

The argument that the delay at synapses is too short for the processes required by the chemical theory can be turned against the electragonists. When an electric current is applied to a nerve, the interval between the instant of shock and the discharge of the nerve impulse ranges between 0.2 and 0.4 msec. The delay between the arrival of the impulse at the superior cervical sympathetic ganglion and the discharge of the ganglion cells is at least 5 times as long as that for the quickest cells and at least 11 times as long for the next quickest (13). If a weak electric current can activate the nerve fibers almost instantly, and if the electric action current activates the ganglionic neurones, as the electragonists argue, why this protracted latent period? The relatively long delay in the ganglion is matched by a similar delay in the motor end plate. Here the electragonists have a real problem. The chemagonists, on the

## MEDIATION OF NERVE IMPULSES

other hand, can readily account for the extra time as due to the requirements of an interposed chemical mediation.

At this point may be mentioned an attempt to bring electrical events in the ganglion into relation to electrical events in the nerve trunk. As shown by Gasser and Erlanger (14) and Graham and Gasser (15) nerves have a heightened electrical excitability during the negative after-potential and a decreased excitability during the subsequent positive after-potential. These two phases, negative and positive, following the spike potential, can also be recorded from the ganglion after it has been stimulated; and Eccles (16, 17) has claimed that submaxi-

mal volleys of nerve impulses delivered to the ganglion during its negative after-potential evoke larger spikes, and when delivered during the positive after-potential evoke smaller spikes, than the original. These observations, if confirmed, might bring support to the electrical theory. In the present discussion a detailed review of the evidence and arguments against Eccles' contentions is impossible. Suffice it to remark that Rosenbluth and Simeone (18) have recorded variations of the spike potentials of the ganglion, responding to maximal and submaximal stimulation, that were precisely opposite to what the inferences of Eccles would predict — maximal spikes when the ganglion was maximally positive, and noteworthy decrease of the spikes at the peak of the negative phase (see Figure 4). Furthermore they noted that by altering conditions the synaptic delay could be shortened to less than 1 msec. or increased to twice its usual length. Because of the lack of correlation between after-potentials and the responsiveness of the ganglion cells, and also because of the large variations in synaptic delay, as well as for other reasons, they concluded that the electrical theory, instead of agreeing with observed facts, is definitely opposed by them.

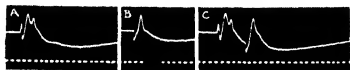


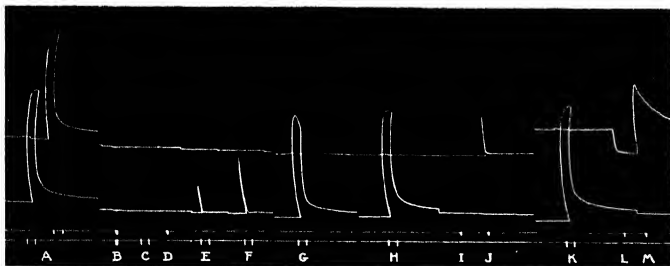
Figure 4. Increase of the response to a test volley when occurring during the positive after-potential of a conditioning volley. Both stimuli delivered through the same electrodes to the preganglionic nerve trunk. A: conditioning volley alone. B: test volley alone. C: test volley applied 52 msec. after the conditioning volley.

The action of curare confronts the electragonists with a serious problem. As is well known, this drug is capable of paralyzing muscles by interrupting the nervous influence at the neuromuscular junctions. It does not, however, prevent the discharge of acetylcholine at the nerve endings; the nervous action currents, therefore, run their full course. And with small doses of curare the inherent irritability of the paralyzed muscle is not altered (19). Yet the muscle does not respond to the nerve impulse. The paralyzing effect of curare at the motor end plate offers no difficulty for the chemical theory. What curare does is to reduce the sensitiveness of the responsive element in the plate. In studies on denervated skeletal muscle Rosenblueth and Luco (20) demonstrated that even with doses too small to stop action of the respiratory muscles curare markedly lessens the contractions resulting from uniform injections of acetylcholine. And the same phenomenon is seen in normally innervated skeletal muscle (21). In short, curare acts as if it raised the threshold to acetylcholine. Then the normal amount of the mediator which is released by the nerve impulse is incapable of causing the muscle to contract. Thus, while a curare paralysis displays a phenomenon hard for the electragonists to explain, it forms a part of the factual support for chemical transmission.

Recent studies on curarized end plates present another hard problem to the electragonists. In experiments performed by Luco and Rosenblueth (22) two symmetrical muscles (e. g., the soleus of each side) were arranged to record simultaneously. After a paralyzing dose of curare had been injected the motor nerve of one of the muscles was stimulated continuously at the rate of 60 shocks per sec. When the control muscle was periodically tested through its nerve, and by its response proved that complete recovery from curarization had occurred, the continuously stimulated muscle was usually not contracting at all. If given a brief rest and then again subjected to the stimulation, it contracted momentarily and thereupon promptly became inert; i. e., it manifested extreme fatigue (see Figure 5).



## MEDIATION OF NERVE IMPULSES



*Figure 5.* Cat. Abdominal sympathetic chains removed 12 days previously. Dial. Adrenals ligated. Gastrocnemius-plantaris (without soleus) recording on both sides. The upper signal marks the beginning or end of stimulation of the right popliteal nerve, which was activated continuously throughout curarization (upper record). The lower signal indicates the occasional tests made on the other, control side (lower record). A, 10:30 A. M.; 5-sec. tests before curare; a first dose of curare was injected at 10:33; it did not produce complete paralysis of the muscles. B, 10:56; additional small dose of curare. C, 10:59; complete curarization. D, 11:00; beginning of prolonged stimulation of right nerve to muscle on the upper record. E, 11:15; F, 11:30; G, 12:30 P. M.; increasing responses of control muscle showing progressive decurarization. H, 12:59; complete decurarization. I, 1:00; prolonged stimulation of right nerve stopped; the absence of a relaxation shows that no tension had developed throughout the period of stimulation. J, 1:01; stimulus reapplied to the right nerve; a comparison of this record with A shows the marked transmission-fatigue consequent to prolonged stimulation. K, 3:56; the control response is equal to that at the end of decurarization (H); the continuously stimulated muscle is developing some tension. L, 3:57; prolonged stimulation of right nerve stopped again. M, 3:58; stimulus reapplied to right nerve.

Clearly this was not a fatigue of the muscle, for the curare throughout the whole period of stimulation had protected the muscle from being forced into activity. And it was not a fatigue of the nerve, for that remarkable structure can carry impulses at the rate of 60 per sec. for an indefinite time. Furthermore, records of the spike potentials proved that the nerve impulses might be full sized when they were having no effect. The fatigue occurred at the point of transmission. Even if curare should somehow block the nerve action current, the electragnosts would have trouble solving this riddle, because the curare block disappeared while the continuous stimulation of the nerve was evoking no muscular contraction. In other words, at a time when both the electrical excitability of the muscle and the

spike potentials of the nerve were quite normal, the nerve impulses failed to stimulate. Before this situation the electrical theory breaks down. The chemical theory, on the other hand, has no trouble in explaining it. Prolonged rapid stimulation results, after an initial large outburst of acetylcholine, in a gradually reduced discharge until the concentration falls below the effective range (23). In these circumstances, according to the chemical theory, a muscular response would not occur. And it does not.

In order to introduce another important series of facts supporting the theory of chemical transmission at muscular and ganglionic synapses, I wish to present evidence regarding the action of eserine. (That term will be used as equivalent to physostigmine and prostigmin.) The pertinent facts may be summarized as follows:

1. Acetylcholine (small amounts) stimulates.
2. Acetylcholine (large amounts) paralyzes.

∴ 2 thresholds :	{ Paralytic— Stimulatory—	Paralysis
		Effective range
		Subthreshold

3. Curare raises both thresholds.
4. Acetylcholine is quickly destroyed by an esterase.
5. Acetylcholine is protected from destruction by eserine (prostigmin).
6. Acetylcholine production is less as stimuli are repeated.

As Rosenblueth and Morison (24) demonstrated, eserine has quite opposite effects on two symmetrical muscles, one recording responses to slowly repeated nerve impulses, the other responding to similar but rapid stimulation. A small intravenous injection of eserine causes the less frequently stimulated muscle to contract more strongly, and simultaneously it causes the corresponding, more frequently stimulated muscle to contract less strongly (see Figure 6). This very striking difference of effect is not due to altered action currents, for the intensity of

## MEDIATION OF NERVE IMPULSES

the nerve spike potential is not affected by eserine. Eccles (25) has suggested that eserine exerts its influence by increasing the sensitiveness of the responding structures. Obviously that suggestion does not apply to the depressant effect on the rapidly stimulated muscle. That effect has its explanation in the change



*Figure 6.* Upper record: right gastrocnemius-soleus stimulated indirectly by maximal shocks at a frequency of 1 per 10 sec. Lower record: left gastrocnemius-soleus stimulated indirectly by maximal shocks at a frequency of 1.3 per sec. At signal: physostigmine, 0.6 mg. per kg. Time signal: 30 sec. intervals.

produced in the functioning of such a muscle when acetylcholine is injected so that it is present in excess; the contractions are depressed as they are depressed by eserine. The action of eserine can therefore reasonably be accounted for, in these circumstances, by its shielding of the acetylcholine, as it is set free by the nerve impulses, against attack by cholinesterase. Since the impulses are quickly recurring, the time for destruction is abbreviated, and with protective eserine at hand the acetylcholine accumulates until it reaches a concentration which paralyzes some of the end plates. Thus would the chemag-onists clear up the phenomenon.

But how account for the opposite effect in the slowly stimulated muscle? That, again, finds an explanation in the shielding of acetylcholine by eserine. Under normal conditions a single volley of nerve impulses evokes a single response in a muscle, as registered by the spike potential of the muscle fibers. If eser-

ine is administered, however, a single volley from the nerve induces a repetitive response and hence a greater muscular shortening (see Figure 7, record 3). As Brown, Dale, and Feldberg (21) have pointed out, this sort of response to single volleys after eserine may reasonably be attributed to the persistence of the acetylcholine released by the motor impulse at the neuromuscular junctions.

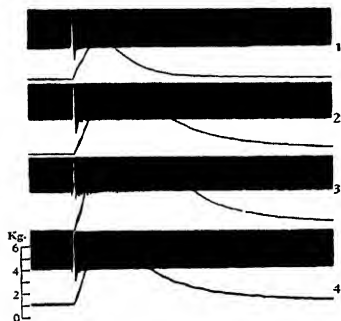


Figure 7. Optical myogram and action potential of gastrocnemius of spinal cat, 2.6 kg. (1) before eserine; (2) 3 min. after 0.8 mg. eserine intravenously; (3) 10 min. later; (4) 2 hrs. later. (From Brown, Dale, and Feldberg, 1936.)

The only difficulty the chemagonists meet in clarifying the remarkably opposed effects of eserine on slowly and on rapidly stimulated muscles is that of understanding why these effects are prolonged. It is known that potassium ions are liberated simultaneously with acetylcholine, and they, as persistent agents, may play a role not yet evident. Whatever may be the final chapter of this story it is clear now that the results of injecting eserine are such as to place still unsurmounted obstacles in the path of the electragonists, while presenting the chemagonists additional support for their views.

Thus far the discussion has been concerned mainly with transmission at neuromuscular synapses. Many years ago Elliott (26) called attention to striking resemblances between characteristics of motor end plates and of cell bodies in sympathetic ganglia. For example, nicotine in small doses stimulates at both places, and in larger doses paralyzes both; curare blocks transmission to skeletal muscle, and in greater concentration has a similar influence in the ganglion; and, whereas the

## MEDIATION OF NERVE IMPULSES

postganglionic axons cannot replace motor fibers, the preganglionic axons, effective on the ganglionic cells, can be cross-sutured and will grow out and make good functional union with striated muscle cells. To these resemblances we can now add the following: acetylcholine is produced at both synapses, neuromuscular and neuroneuronal; acetylcholine in small doses excites at both synapses; and when present in excess acetylcholine, like nicotine, has a depressant or paralyzing influence. We have learned recently that there are still other remarkable similarities between these two transmission points. It will be convenient, therefore, to consider them together.

As shown by Rosenblueth and Morison (24), if a muscle is stimulated through its nerve at a rapid rate—e. g., 540 shocks per sec.—the muscle first shortens in a sharp strong contraction, then changes promptly to a low contraction or fails to contract at all, and thereupon it shortens in a strong contraction again (see Figure 8). Thus with recurring impulses of a constant high frequency the muscular response manifests three distinct stages: a sequence of plus, minus, plus. Now if the stimulation is continued the muscle enters a well-known fourth stage, that of lessened efficiency or fatigue. If the stimulation is at a slow rate from the start, stages two and three drop out; then stage one merges directly with stage four. Curiously enough, as revealed in the experiments of Rosenblueth and Luco (27), continuance of the stimulation at a slow fatiguing rate—60 per sec.—causes the muscle to enter gradually a fifth stage, in which its performance progressively improves; the contractions increase in strength, displaying a tension as great as 60 per cent of the highest tension developed at the beginning, and this betterment may last for several hours (see Figure 9).

By use of the nictitating membrane of the cat as an indicator of the discharge from the superior cervical sympathetic ganglion the effect of stimulation of preganglionic fibers on the ganglion cells can be examined. In experiments which Rosenblueth and I (28) performed we were able to duplicate, with

# WALTER B. CANNON

rapidly repeated stimuli, the first four stages—plus, minus, plus, minus—seen in skeletal muscle (see Figure 10). The only difference was a preliminary dose of eserine in experiments on the ganglion. And a few months ago Lanari and Rosenblueth (29) succeeded in recording in the ganglionic performance the fifth stage, that of partial recovery of the original efficiency, if only the fatiguing stimulation is patiently continued (see Figure 11).

Here is a challenging sequence of events. Why do muscle cells and ganglion cells, when continuously stimulated, respond well at first, then poorly; well again, then poorly again; and finally well for a long period? As proved by Rosenblueth and

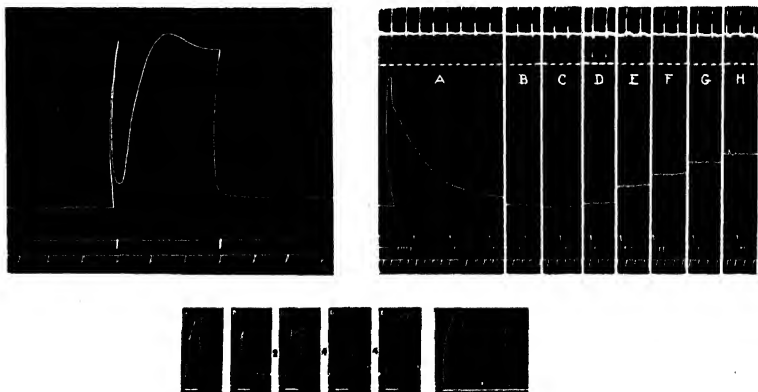
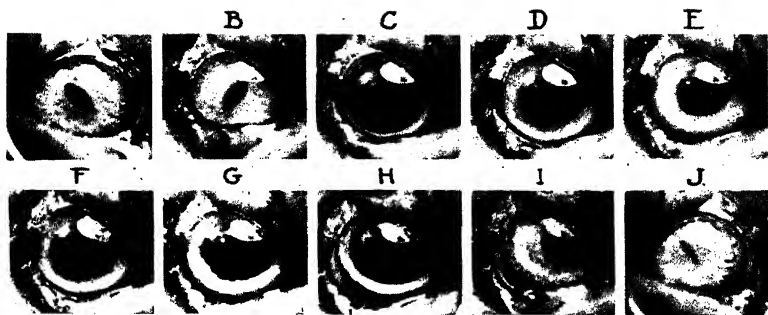


Figure 8 (left). Cat's gastrocnemius-soleus. Popliteal nerve stimulated; frequency 540 per sec.

Figure 9 (right). Electric responses of popliteal nerve and mechanical response of gastrocnemius-plantaris muscles to continuous maximal stimulation at 60 per sec. The nerve spike-potentials were recorded during the corresponding kymograph record strips. A: beginning of stimulation. B to H: 30, 60, 110, 155, 180, 210, and 240 min. later, respectively. Time signals: mechanogram, 1 min.; electrogram, 10 msec. The two upper signals in the kymograph records merely indicate times at which electrical records were taken; they are to be disregarded.

Figure 10 (center). Effects of prostigmin on recurrent maximal tetanic stimulation of preganglionic fibers: A: response of the nictitating membrane after atropine (4 mg.) but before prostigmin; B: 3 min. after prostigmin (2 mg.); intervals between records indicated in minutes (secondary coil at 8 cm.). At extreme right: effect of prostigmin on the late stage of tetanic stimulation of preganglionic fibers. At signal mark prostigmin (1 mg.) injected intravenously.



*Figure 11.* The fifth stage in the cat's iris. A and J: left pupil at beginning and end of the experiment. B: right pupil before stimulation. C: 15 sec. after the beginning of stimulation of the right cervical sympathetic (frequency, 60 per sec.). D to H: 2, 17, 25, 30, and 45 min. after the beginning of stimulation. I: 1 min. after the stop of stimulation.





## MEDIATION OF NERVE IMPULSES

Luco (27), who took records of the nerve action potentials periodically during four or five hours of stimulation, there may be a slow diminution in the height of the spikes or there may be no diminution, but in no case is there any correlation between the electrical phenomena of the nerve and the various stages of muscular activity. These conditions at motor end plates and at ganglionic synapses — and with normal blood supply, let it be noted — disclose a puzzle which the philosophy of the electragonists has not dreamed of. What clue to its solution can the chemagonists present?

As previously remarked, when a nerve is continuously stimulated there is at first a flush discharge of acetylcholine at the nerve endings, and then the discharge gradually diminishes. Now if the stimuli are repeated at very high frequency time will be lacking for destruction of the relatively abundant acetylcholine by cholinesterase; thereby the initial great outburst becomes accentuated and the acetylcholine accumulates until it is present in a paralyzing concentration. Thus the primary contraction of the muscle (stage one) is quickly followed by less contraction or an actual failure to contract. In the ganglion a small dose of eserine, protective against the destructive cholinesterase, allows the acetylcholine to act in the same manner, so that it has the same paralyzing effect. As the primary outburst quickly subsides the depressant concentration of acetylcholine ebbs away, to be replaced by an optimal lower concentration which is highly stimulative; hence the rise of the muscular tension into stage three. That this is the correct explanation is proved by the injection of a small amount of acetylcholine during the depression, whereupon the depressed condition is prolonged; or during stages one or three, whereupon not stimulation but, instead, depression is induced (see Figure 12). The fourth stage, the stage of fatigue, can be explained by so great a reduction in the output of acetylcholine that it fails to stimulate some of the cells; hence the lessened degree of response — i. e., the appearance of "fatigue." That this interpretation is correct is proved by injecting at this stage the same dose of

## WALTER B. CANNON

acetylcholine as that which caused depression in the early stages and finding that it now causes an improvement of the response. This, of course, is strictly in accord with the theory of chemical transmission.

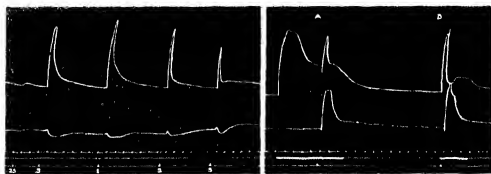
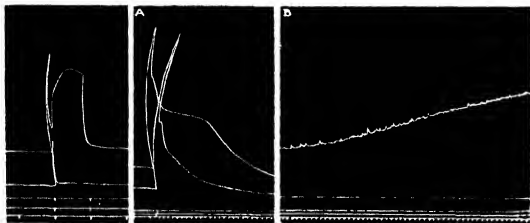


Figure 12. At left: effects of increasing doses of acetylcholine on responses of nictitating membranes; upper record, superior cervical ganglion present; lower, absent; doses (intravenous) in milligrams; time intervals, half-minutes; dial, atropine, adrenals, out. At right: effects of acetylcholine at different phases of tetanic stimulation of preganglionic fibers; upper record, tetanic, and lower record, single shocks, applied to preganglionic fibers; at signal-marks acetylcholine (1 mg.) injected intravenously.

The astonishing recovery of function in the fifth stage has been investigated by Rosenblueth, Lissák, and Lanari (30). They have assayed the acetylcholine content of nerves, quickly frozen *in situ* by carbon-dioxide snow, or instantly excised and dropped into liquid air (in order to prevent any change), after continuous, rapid stimulation for various lengths of time. They found a progressive decrease in the content during the first ten minutes, a decrease which turned to an increase when there was a brief period of rest. In the fifth stage the concentration of acetylcholine had gradually increased over that in the fourth stage. If, now, it is assumed that depolarization of the surface membrane of the nerve fibers is attended by a greater permeability for the acetylcholine contained within them, the output per impulse would vary directly with the concentration. In the fifth stage, when the fibers have more acetylcholine, they give forth more, and the result is a larger response. This is the view presented by Rosenblueth, Lissák, and Lanari. It is substantially based on the agreement between the evidence of acetylcholine liberation from nerves at the different stages of stimulation and the evidence of concentration of acetylcholine in nerve fibers.

## MEDIATION OF NERVE IMPULSES

I have described a group of events in the transmission of nerve impulses from neurone to neurone and from neurone to muscle cell. I can find no adequate explanation for these perplexing events in terms of the electrical theory. On the contrary, I can see that supporters of the theory of chemical transmission need only use demonstrated facts in order to give a reasonable answer to the whole intricate riddle.



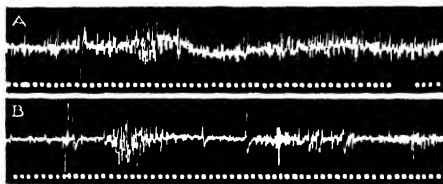
*Figure 13 (left).* Disappearance of the second and third stages of neuromuscular transmission during Wallerian degeneration; Achilles-tendon muscles; indirect stimulation for 30 sec. at 400 per sec.; upper tracing, control; lower tracing, sciatic cut 48 hrs.

*Figure 14 (right).* Disappearance of the fifth stage during Wallerian degeneration; records as in Figure 13; sciatic cut 30 hrs. (lower tracing); continuous stimulation at 60 per sec.; A: first 25 min. of stimulation; B: 75 min. after A.

The main assumption in the foregoing account of the five stages of transmission at peripheral synapses was that the output of acetylcholine per nerve impulse would vary with its concentration in the nerve fiber. The recent research by Rosenblueth, Lissák, and Lanari (30), already noted, has furnished evidence that when a nerve has been cut and is undergoing degeneration there is a gradual reduction of the acetylcholine content after the first twenty-four hours. A highly significant fact is that during these first twenty-four hours no functional change is observable. As the concentration of acetylcholine diminishes in the second twenty-four hours, however, the possibility of evoking stage five is regularly lost, and stages two and three may largely disappear or be wholly absent (see Figures 13 and 14). At the same time the nerve spike potentials are of normal intensity. The results are quite in accord with the assumption mentioned above, for, as will be recalled, tests re-

## WALTER B. CANNON

vealed that stages two and three were due to an abundance of acetylcholine and stage five to an improved production of it. In the degenerative process, with the acetylcholine of the nerve gradually becoming less, these periods of excessive production fade out, and while they do so the attendant variations of activity likewise fade out. And, as might have been anticipated,



*Figure 15.* The Philipeaux-Vulpian phenomenon in response to single nerve volleys (after eserine). Right hypoglossal cut 7 days previously. Dial; atropine (1 mg. per kg.); eserine (1 mg. per kg.); left hypoglossal cut. Single discharges from a photoelectric cell briefly illuminated applied to the right lingual nerve, as shown by the large diphasic stimulus artifact. Time (short intervals), 10 msec. A and B: two different positions of the concentric electrodes in the right side of the tongue.

the fourth stage, that of fatigue, not only outlasts stages two, three, and five, but becomes unduly emphasized; the transmission from degenerating nerves “fatigues” much sooner than normally. Now the further important fact emerges that while nerves in the late steps of degeneration can still conduct, as revealed by a definite, though reduced, spike potential, they may not stimulate across the synapse. All these observations harmonize perfectly with the chemical theory of transmission and find no illumination whatever in the electrical theory.

Perhaps no further testimony is needed to justify the claims of the chemagonists. There is one other observation, however, which is demonstrative and well worth presenting. As is now generally recognized, autonomic nerves do not end on striated muscle. They do end on blood vessels—vessels intimately distributed throughout striated muscle fibers. Among the autonomic nerves which give off acetylcholine is the lingual, supplied to the vessels of the tongue. In 1937 Rosenblueth and Luco (20) recorded the action potentials of the tongue muscles

## MEDIATION OF NERVE IMPULSES

which had been deprived of motor nerves for about a week. Stimulation of the lingual nerve induced the appearance of typical action currents in the completely denervated muscle fibers (see Figure 15). Here is a stimulation of muscular units clearly not due to electrical transmission, for the nerve stimulated was not distributed to the muscle which responded; and furthermore the latent period (60 msec.) was too long for that occurrence. It was a clear instance of a contractile response to acetylcholine, liberated by nerve impulses, to be sure, but not nerve impulses delivered to the contracting fibers. Electrical transmission was definitely excluded. Until the electragonists can display an instance of electrical transmission without acetylcholine at neuromuscular and neuroneuronal synapses their argument cannot be on the same footing as that of the chemagonists.

Meanwhile they have arduous tasks in explaining away the evidence which has repeatedly proved the adequacy of the chemical theory of nervous transmission.

### *References*

1. Gasser, H. S. 1937. Harvey Lectures 32:169-74.
2. Schmitt, F. O., and Gasser, H. S. 1933. Am. Jour. Physiol. 104:320.
3. Hodgkin, A. L. 1937. Jour. Physiol. 90:183, 211.
4. Blair, E. A., and Erlanger, J. 1939. Am. Jour. Physiol. 126:97.
5. Bronk, D. W., and Brink, F. 1939. Ann. Rev. Physiol. 1:398.
6. Forbes, A. 1939. Jour. Neurophysiol. 2:465.
7. Rosenblueth, A., and Cannon, W. B. 1934. Am. Jour. Physiol. 108:884.
8. Cannon, W. B., and Rosenblueth, A. 1937. Compt. rend. de soc. de biol. 124:1262.
9. Marnay, A., and Nachmansohn, D. 1937. Compt. rend. de soc. de biol. 125:41.
10. Feng, T. P., and Ting, Y. C. 1938. Chinese Jour. Physiol. 13:141.
11. Eccles, J. C. 1937. Physiol. Rev. 17:538.
12. Brown, G. L. 1937. Physiol. Rev. 17:485.
13. Eccles, J. C. 1935. Jour. Physiol. 85:179.
14. Gasser, H. S., and Erlanger, J. 1930. Am. Jour. Physiol. 94:247.
15. Graham, H. T., and Gasser, H. S. 1934. Proc. Soc. Exper. Biol. Med. 32:553.
16. Eccles, J. C. 1935. Jour. Physiol. 85:464.
17. ——— 1936. Jour. Physiol. 88:1.
18. Rosenblueth, A., and Simeone, F. A. 1938. Am. Jour. Physiol. 122:688.

WALTER B. CANNON

19. Grundfest, H. 1932. Jour. Physiol. 76:95.
20. Rosenblueth, A., and Luco, J. V. 1937. Am. Jour. Physiol. 120:781.
21. Brown, G. L., Dale, H. H., and Feldberg, W. 1936. Jour. Physiol. 87:394.
22. Luco, J. V., and Rosenblueth, A. 1939. Am. Jour. Physiol. 126:58.
23. Brown, G. L., and Feldberg, W. 1936. Jour. Physiol. 88:265.
24. Rosenblueth, A., and Morison, R. S. 1937. Am. Jour. Physiol. 119:236.
25. Eccles, J. C. 1936. Ergebn. d. Physiol. 38:339.
26. Elliott, T. R. 1907. Jour. Physiol. 35:367.
27. Rosenblueth, A., and Luco, J. V. 1939. Am. Jour. Physiol. 126:39.
28. Cannon, W. B., and Rosenblueth, A. 1937. Am. Jour. Physiol. 119:221.
29. Lanari, A., and Rosenblueth, A. 1939. Am. Jour. Physiol. 127:347.
30. Rosenblueth, A., Lissák, K., and Lanari, A. In press. Am. Jour. Physiol.

